

Please visit our web site (<http://genome.rtc.riken.go.jp>) for further details.

FEATURES  
SOURCE

```
Location/Qualifiers
1. .259
  /organism="Mus musculus"
  /mol_type="rRNA"
  /db_xref="taxon:10090"
  /clone="B230206M01"
  /sex="male"
  /tissue_type="corpora qu
  /dev_stages="adult"
  /lab_host="DH10B"
  /clone_lib="RIKEN full-l
  /cloning_center="quadrigemina"
```

```

/clone_lib="RIKEN full-length enriched, adult male corpora
quadrigemina"

```

quadrigemina  
/note=Site\_1: SalI; Site\_2: BamHI; cDNA library was  
prepared and sequenced in Mouse Genome Encyclopedia  
Project of Genome Exploration Research Group in Riken  
Institute, Ibaraki, Japan.

RIKEN. Division of  
contributed to pr  
primed with a pr

Preparatory using standard use of the advanced full length by creating a template cDNA through one round of normalization to R60 and subtraction to R60. 459.0 Second strand cDNA was prepared with the primer adapter of sequence [5', GAGACGAGATTCGAGTAAATTAATTCACCCCCCCCC 3']. cDNA was cleaved with XhoI and BamHI. Vector: a modified pLuciferase KS(+) after bulk excision from Lambda PLC I."

Query Match 84.0%; Score 16.8; DB 10; Length 259;  
Best Local Similarity 90.0%; Pred. No. 2.2e+03;  
Matches 18; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY	1	CTGGACAGGTTAGGCTTTG	20
Db	72	CTTGCACAGTTATGGCTTTG	91

Search completed: February 20, 2004, 01:41:36  
Job time : 137.422 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: February 19, 2004, 21:32:00 ; Search time 3.2454 Seconds  
(without alignments)  
3419.919 Million cell updates/sec

Title: US-09-939-853a-140

Perfect score: 20

Sequence: 1 ctggacaggttaggcttgg 20

Scoring table: IDENTITY NUC

Gapop 10.0 , Gapext 1.0

Searched: 682709 seqs, 277475446 residues

Total number of hits satisfying chosen parameters: 1365418

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

Issued Patents NA.\*

1: /cgn2\_6/ptodata/2/ina/5A COMB.seq.\*

2: /cgn2\_6/ptodata/2/ina/5B COMB.seq.\*

3: /cgn2\_6/ptodata/2/ina/6A COMB.seq.\*

4: /cgn2\_6/ptodata/2/ina/6B COMB.seq.\*

5: /cgn2\_6/ptodata/2/ina/PTCUS COMB.seq.\*

6: /cgn2\_6/ptodata/2/ina/backfiles1.seq.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	15.8	79.0	99916	4	US-09-816-095-3
2	15.2	76.0	303	4	US-09-489-039A-4299
3	15.2	76.0	1515	4	US-09-071-035-431
4	15.2	76.0	1803	4	US-09-071-035-429
5	15.2	76.0	2481	4	US-09-134-000C-3193
6	15.2	76.0	2611	4	US-09-620-312D-925
7	15.2	76.0	3614	4	US-09-221-013A-9
8	15.2	76.0	48974	3	US-08-920-422-17
9	14.8	74.0	514	4	US-09-621-976-14354
10	14.8	74.0	2068	2	US-08-466-589-1
11	14.8	74.0	2068	3	US-08-700-636-1
12	14.8	74.0	2068	3	US-08-467-574-1
13	14.8	74.0	2068	4	US-09-217-345-1
14	14.8	74.0	2068	4	US-09-892-985-1
15	14.8	74.0	2277	1	US-08-436-855A-1
16	14.8	74.0	2277	4	US-08-487-596-1
17	14.8	74.0	2352	2	US-08-899-909A-21
18	14.8	74.0	2352	4	US-09-156-163A-21
19	14.8	74.0	2352	4	US-09-982-308B-21
20	14.8	74.0	2430	1	US-09-062-368-1
21	14.8	74.0	2664	4	US-08-660-451A-1
22	14.8	74.0	6268	4	US-09-566-921-57
23	14.8	74.0	18994	1	US-08-459-386-4
24	14.8	74.0	18994	2	US-08-282-696-4
25	14.4	72.0	34063	4	US-09-453-702B-96
26	14.4	72.0	36519	3	US-08-923-137-2
27	14.2	71.0	998	4	US-09-671-317-191

28	14.2	71.0	1299	4	US-09-404-641-80	Sequence 80, Appl
29	14.2	71.0	1298	4	US-09-404-641-69	Sequence 69, Appl
30	14.2	71.0	1398	4	US-09-328-352-2591	Sequence 2591, Ap
31	14.2	71.0	1572	4	US-09-620-312D-886	Sequence 886, App
32	14.2	71.0	1735	4	US-09-404-641-84	Sequence 84, Appl
33	14.2	71.0	1773	4	US-09-134-000C-1350	Sequence 1350, Ap
34	14.2	71.0	1799	4	US-09-732-234-5	Sequence 5, Appl
35	14.2	71.0	1799	4	US-09-784-859-5	Sequence 5, Appl
36	14.2	71.0	3588	4	US-09-620-312D-863	Sequence 863, App
37	14.2	71.0	41171	4	US-08-311-731A-122	Sequence 122, App
38	14.2	71.0	1664976	4	US-08-916-421B-1	Sequence 1, Appl
39	14.2	71.0	1664976	4	US-08-916-421B-1	Sequence 1, Appl
40	14	70.0	1326	4	US-09-328-352-3788	Sequence 3788, Ap
41	13.8	69.0	159	3	US-08-651-136C-47	Sequence 47, Appl
42	13.8	69.0	159	4	US-09-229-911A-47	Sequence 47, Appl
43	13.8	69.0	180	3	US-08-651-136C-69	Sequence 69, Appl
44	13.8	69.0	180	4	US-09-229-911A-69	Sequence 69, Appl
45	13.8	69.0	291	4	US-09-621-976-17092	Sequence 17092, A

#### ALIGNMENTS

RESULT 1  
US-09-816-095-3  
; Sequence 3, Application US/09816095  
; Patent No. 6654084  
; GENERAL INFORMATION:  
; APPLICANT: GAN, Weiniu  
; TITLE OF INVENTION: ISOLATED HUMAN ENZYME PROTEINS, NUCLEIC  
; TITLE OF INVENTION: ACID MOLECULES ENCODING HUMAN ENZYME PROTEINS, AND USES  
; TITLE OF INVENTION: THEROOF  
; FILE REFERENCE: CLO01147  
; CURRENT APPLICATION NUMBER: US/09/816,095  
; CURRENT FILING DATE: 2001-03-26  
; NUMBER OF SEQ ID NOS: 5  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 3  
; LENGTH: 99916  
; TYPE: DNA  
; ORGANISM: Human  
; FEATURE:  
; NAME/KEY: misc feature  
; LOCATION: (1)..(99916)  
; OTHER INFORMATION: n = A,T,C or G  
US-09-816-095-3

Query Match 79.0%; Score 15.8; DB 4; Length 99916;  
Best Local Similarity 89.5%; Pred. No. 86;  
Matches 17; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 2 TGGACAGGTAGGCTTTG 20  
DB 5532 TGGACAGATTAGGCTTTG 5550

#### RESULT 2

US-09-489-039A-4299/c  
; Sequence 4299, Application US/09489039A  
; Patent No. 6610836  
; GENERAL INFORMATION:  
; APPLICANT: Gary Breton et. al  
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO KLEBSIELLA  
; TITLE OF INVENTION: PNEUMONIAE FOR DIAGNOSTICS AND THERAPEUTICS  
; FILE REFERENCE: 2709-2004001  
; CURRENT APPLICATION NUMBER: US/09/489,039A  
; CURRENT FILING DATE: 2000-01-27  
; PRIOR APPLICATION NUMBER: US 60/117,747  
; PRIOR FILING DATE: 1999-01-29  
; NUMBER OF SEQ ID NOS: 14342  
; SEQ ID NO 4299  
; LENGTH: 303  
; TYPE: DNA



```
; Sequence 925, Application US/09620312D
; Patent No. 6569662
; GENERAL INFORMATION:
; APPLICANT: Tang, Y. Tom
; APPLICANT: Liu, Chenchua
; APPLICANT: Asundi, Vinod
; APPLICANT: Zhang, Jie
; APPLICANT: Ren, Feiyan
; APPLICANT: Chen, Rui-hong
; APPLICANT: Zhao, Qing A.
; APPLICANT: Wehrman, Tom
; APPLICANT: Xue, Aidong J.
; APPLICANT: Yang, Yonghong
; APPLICANT: Wang, Jian-Rui
; APPLICANT: Zhou, Ping
; APPLICANT: Ma, Yunqing
; APPLICANT: Wang, Dunrui
; APPLICANT: Wang, Zhiwei
; APPLICANT: John Tillingshast
; APPLICANT: Dmanac, Radoje T.
; TITLE OF INVENTION: Polypeptides
; TITLE OF INVENTION: Nucleic Acids and
; FILE REFERENCE: 784CIP2B
; CURRENT APPLICATION NUMBER: US/09/620,312D
; CURRENT FILING DATE: 2000-07-19
; PRIOR APPLICATION NUMBER: 09/552,317
; PRIOR FILING DATE: 2000-04-25
; PRIOR APPLICATION NUMBER: 09/488,725
; PRIOR FILING DATE: 2000-01-21
; NUMBER OF SEQ ID NOS: 1105
; SOFTWARE: P1_FL_Genes Version 1.0
; SEQ ID NO 925
; LENGTH: 2611
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (290)..(1885)
US-09-620-312D-925

Query Match      76.0%; Score 15.2; DB 4; Length 2611;
Best Local Similarity 85.0%; Pred. No. 98;
Matches 17; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 CTGACAGGTTAGGCTTTG 20
Db 789 CTGAACAGATTAGTCTTTG 808

RESULT 7
US-09-221-013A-9/c
; Sequence 9, Application US/09221013A
; Patent No. 6495740
; GENERAL INFORMATION:
; APPLICANT: Arioli, Antonio
; APPLICANT: Williamson, Richard E.
; APPLICANT: Betzner, Andreas S.
; APPLICANT: Peng, Liangcai
; TITLE OF INVENTION: Manipulation of cellulose and/or Beta-1,4-glucan
; FILE REFERENCE: 96-98
; CURRENT APPLICATION NUMBER: US/09/221,013A
; CURRENT FILING DATE: 1998-12-23
; PRIOR APPLICATION NUMBER: PCT/AU97/00402
; PRIOR FILING DATE: 1997-06-24
; PRIOR APPLICATION NUMBER: AU P00699
; PRIOR FILING DATE: 1996-06-27
; NUMBER OF SEQ ID NOS: 37
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 9
; LENGTH: 3614
; TYPE: DNA
; ORGANISM: Arabidopsis thaliana
; FEATURE:

; NAME/KEY: CDS
; LOCATION: (217)..(3411)
US-09-221-013A-9

Query Match      76.0%; Score 15.2; DB 4; Length 3614;
Best Local Similarity 85.0%; Pred. No. 1e+02;
Matches 17; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 CTGACAGGTTAGGCTTTG 20
Db 1600 CAGGACATTTTAGGGCTTTG 1581

RESULT 8
US-08-920-422-17/c
; Sequence 17, Application US/08920422A
; Patent No. 6255473
; GENERAL INFORMATION:
; APPLICANT: Vitek, Michael P.
; APPLICANT: Mitsuda, No. 6255473iaki
; APPLICANT: Roses, Allen D.
; TITLE OF INVENTION: Presenilin-1 Gene Promoter
; FILE REFERENCE: VITEKPRESENTINILIN
; CURRENT APPLICATION NUMBER: US/08/920,422A
; CURRENT FILING DATE: 1997-08-29
; NUMBER OF SEQ ID NOS: 22
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 17
; LENGTH: 48974
; TYPE: DNA
; ORGANISM: Mus musculus
US-08-920-422-17

Query Match      76.0%; Score 15.2; DB 3; Length 48974;
Best Local Similarity 85.0%; Pred. No. 1.6e+02;
Matches 17; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 CTGACAGGTTAGGCTTTG 20
Db 33769 CTGGCCAGGATAGGGCTGTG 33750

RESULT 9
US-09-621-976-14354/c
; Sequence 14354, Application US/09621976
; Patent No. 6639063
; GENERAL INFORMATION:
; APPLICANT: Dumas Milne Edwards, J.B.
; APPLICANT: Jobert, S.
; APPLICANT: Giordano, J.Y.
; TITLE OF INVENTION: ESTs and Encoded Human Proteins.
; FILE REFERENCE: GENSET.054PR2
; CURRENT APPLICATION NUMBER: US/09/621,976
; CURRENT FILING DATE: 2000-07-21
; NUMBER OF SEQ ID NOS: 19335
; SOFTWARE: Patent.pm
; SEQ ID NO 14354
; LENGTH: 514
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: 254
; OTHER INFORMATION: n=a, g, c or t
US-09-621-976-14354

Query Match      74.0%; Score 14.8; DB 4; Length 514;
Best Local Similarity 88.9%; Pred. No. 1.2e+02;
Matches 16; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 3 GGACAGGTTAGGCTTTG 20
Db 311 GGACAGGTTAGGCTTTG 294
```

```

US-08-466-589-1/c
; Sequence 1, Application US/08466589
; Patent No. 5837489
; GENERAL INFORMATION:
; APPLICANT: Elliot, Kathryn J.
; APPLICANT: Ellis, Steven B.
; APPLICANT: Harpold, Michael M.
; TITLE OF INVENTION: HUMAN NEURONAL NICOTINIC ACETYLCHOLINE
; TITLE OF INVENTION: RECEPTOR COMPOSITIONS AND METHODS EMPLOYING SAME
; NUMBER OF SEQUENCES: 12
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Brown, Martin, Haller & McClain
; STREET: 1660 Union Street
; CITY: San Diego
; STATE: CA
; COUNTRY: USA
; ZIP: 92101-2926
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; OPERATING SYSTEM: DOS
; SOFTWARE: FastSeq Version 1.5
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/466,589
; FILING DATE: June 5, 1995
; CLASSIFICATION: 536
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/028,031
; FILING DATE: March 8, 1993
; ATTORNEY/AGENT INFORMATION:
; NAME: Seidman, Stephanie L
; REGISTRATION NUMBER: 33,779
; REFERENCE/DOCKET NUMBER: 6362-9950
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 619-238-0999
; TELEFAX: 619-238-0062
; TELEX:
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2068 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: both
; TOPOLOGY: both
; MOLECULE TYPE: cdna
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 166..1752
US-08-466-589-1

Query Match 74.0%; Score 14.8; DB 2; Length 2068;
Best Local Similarity 88.9%; Pred. No. 1.5e+02;
Matches 16; Conservative 0; Mismatches 2; Indels 0; Gaps

QY 3 GGACGAGTTAGGGCTTTG 20
DB 144 GGTGAGGTGAGGCTTTG 127

RESULT 11
US-08-700-636-1/c
; Sequence 1, Application US/08700636
; Patent No. 5910582
; GENERAL INFORMATION:
; APPLICANT: Elliot, Kathryn J.
; APPLICANT: Ellis, Steven B.
; APPLICANT: Harpold, Michael M.
; TITLE OF INVENTION: HUMAN NEURONAL NICOTINIC ACETYLCHOLINE
; TITLE OF INVENTION: RECEPTOR COMPOSITIONS AND METHODS EMPLOYING SAME
; NUMBER OF SEQUENCES: 12
; CORRESPONDENCE ADDRESS:

```

```

; ADDRESSEE: Pretty, Schroeder, Brueggemann & Clark
; STREET: 444 South Flower Street, Suite 2000
; CITY: Los Angeles
; STATE: CA
; COUNTRY: USA
; ZIP: 90071
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; OPERATING SYSTEM: IBM PC compatible
; SOFTWARE: PatentIn Release: PC-DOS/MS-DOS
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/700,636
; FILING DATE: 16-JUL-1996
; CLASSIFICATION: 536
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/028,031
; FILING DATE: 08-MAR-1993
; ATTORNEY/AGENT INFORMATION:
; NAME: Reiter, Stephen E.
; REGISTRATION NUMBER: 31,192
; REFERENCE/DOCKET NUMBER: P41 9368
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 619-546-4737
; TELEFAX: 619-546-9392
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2068 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: both
; TOPOLOGY: both
; MOLECULE TYPE: cDNA
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 166..1752
;
US-08-700-636-1

```

;; PRIOR APPLICATION DATA:  
;; APPLICATION NUMBER: US 08/028,031  
;; FILING DATE: March 8, 1993  
;; ATTORNEY/AGENT INFORMATION:  
;; NAME: Seidman, Stephanie L  
;; REGISTRATION NUMBER: 33,779  
;; REFERENCE/DOCKET NUMBER: 6362-9949  
;; TELEPHONE: 619-238-0999  
;; TELEFAX: 619-238-0062  
;; TELEX:  
;; INFORMATION FOR SEQ ID NO: 1:  
;; SEQUENCE CHARACTERISTICS:  
;; LENGTH: 2068 base pairs  
;; TYPE: nucleic acid  
;; STRANDEDNESS: both  
;; TOPOLOGY: both  
;; MOLECULE TYPE: cDNA  
;; FEATURE:  
;; NAME/KEY: CDS  
;; LOCATION: 166..1752  
US-08-467-574-1

Query Match 74.0%; Score 14.8; DB 3; Length 2068;  
Best Local Similarity 88.9%; Pred. No. 1.5e+02;  
Matches 16; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 3 GGACAGGTTAGGGCTTTG 20  
Db 144 GGTGAGTCAGGGCTTTG 127

RESULT 13  
US-09-217-345-1/c  
; Sequence 1, Application US/09217345  
; Patent No. 6303753  
; GENERAL INFORMATION:  
; APPLICANT: Elliot, Kathryn J.  
; APPLICANT: Ellis, Steven B.  
; APPLICANT: Harpold, Michael M.  
; TITLE OF INVENTION: HUMAN NEURONAL NICOTINIC ACETYLCHOLINE  
; TITLE OF INVENTION: RECEPTOR COMPOSITIONS AND METHODS EMPLOYING SAME  
; NUMBER OF SEQUENCES: 12  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Heller Ehrman White & McAuliffe  
; STREET: 4250 Executive Square, 7th Floor  
; CITY: La Jolla  
; STATE: CA  
; COUNTRY: USA  
; ZIP: 92037  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Diskette  
; COMPUTER: IBM Compatible  
; OPERATING SYSTEM: DOS  
; SOFTWARE: FastSeq Version 1.5  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/09/217,345  
; FILING DATE: 21-DEC-98  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 08/467,574  
; FILING DATE: 05-JUN-95  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 08/466,589,  
; FILING DATE: 05-JUN-95  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 08/028,031  
; FILING DATE: 08-MAR-93  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Seidman, Stephanie L  
; REGISTRATION NUMBER: 33,779  
; REFERENCE/DOCKET NUMBER: 24735-9949B  
; TELEPHONE: 619-450-8400  
; TELEFAX: 619-450-8400

;; TELEFAX: 619-587-5360  
;; TELEX:  
;; INFORMATION FOR SEQ ID NO: 1:  
;; SEQUENCE CHARACTERISTICS:  
;; LENGTH: 2068 base pairs  
;; TYPE: nucleic acid  
;; STRANDEDNESS: both  
;; TOPOLOGY: both  
;; MOLECULE TYPE: cDNA  
;; FEATURE:  
;; NAME/KEY: CDS  
;; LOCATION: 166..1752  
US-09-217-345-1

Query Match 74.0%; Score 14.8; DB 4; Length 2068;  
Best Local Similarity 88.9%; Pred. No. 1.5e+02;  
Matches 16; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 3 GGACAGGTTAGGGCTTTG 20  
Db 144 GGTGAGTCAGGGCTTTG 127

RESULT 14  
US-09-892-985-1/c  
; Sequence 1, Application US/09892985  
; Patent No. 6664375  
; GENERAL INFORMATION:  
; APPLICANT: Elliot, Kathryn J.  
; APPLICANT: Ellis, Steven B.  
; APPLICANT: Harpold, Michael M.  
; TITLE OF INVENTION: HUMAN NEURONAL NICOTINIC ACETYLCHOLINE  
; TITLE OF INVENTION: RECEPTOR COMPOSITIONS AND METHODS EMPLOYING SAME  
; NUMBER OF SEQUENCES: 12  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Heller Ehrman White & McAuliffe  
; STREET: 4250 Executive Square, 7th Floor  
; CITY: La Jolla  
; STATE: CA  
; COUNTRY: USA  
; ZIP: 92037  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Diskette  
; COMPUTER: IBM Compatible  
; OPERATING SYSTEM: DOS  
; SOFTWARE: FastSeq Version 1.5  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/09/892,985  
; FILING DATE: 27-Jun-2001  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US/09/217,345  
; FILING DATE: 21-DEC-98  
; APPLICATION NUMBER: US 08/467,574  
; FILING DATE: 05-JUN-95  
; APPLICATION NUMBER: US 08/466,589,  
; FILING DATE: 05-JUN-95  
; APPLICATION NUMBER: US 08/028,031  
; FILING DATE: 08-MAR-93  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Seidman, Stephanie L  
; REGISTRATION NUMBER: 33,779  
; REFERENCE/DOCKET NUMBER: 24735-9949B  
; TELEPHONE: 619-450-8400  
; TELEFAX: 619-587-5360  
; TELEX: <Unknown>

;; INFORMATION FOR SEQ ID NO: 1:  
;; SEQUENCE CHARACTERISTICS:  
;; LENGTH: 2068 base pairs  
;; TYPE: nucleic acid  
;; STRANDEDNESS: both  
;; TOPOLOGY: both  
;; MOLECULE TYPE: cDNA

Search completed: February 20, 2004, 01:45:09  
Job time : 6.2454 secs

FEATURE:  
NAME/KEY: CDS  
LOCATION: 166..1752  
SEQUENCE DESCRIPTION: SEQ ID NO: 1:  
US-09-892-985-1  
Query Match 74.0%; Score 14.8; DB 4; Length 2068;  
Best Local Similarity 88.9%; Pred. No. 1.5e+02;  
Matches 16; Conservative 0; Mismatches 2; Indels 0; Gaps 0;  
QY 3 GGACAGGTAGGGCTTTG 20  
DB 144 GGTCAAGTCAGGGCTTTG 127

RESULT 15  
US-08-496-855A-1/C  
; Sequence 1, Application US/08496855A  
; Patent No. 5801232  
; GENERAL INFORMATION:  
; APPLICANT: Elliot, Kathryn J.  
; APPLICANT: Ellis, Steven B.  
; APPLICANT: Harpold, Michael M.  
; TITLE OF INVENTION: HUMAN NEURONAL NICOTINIC ACETYLCHOLINE  
; TITLE OF INVENTION: RECEPTOR COMPOSITIONS AND METHODS EMPLOYING SAME  
; NUMBER OF SEQUENCES: 6  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Brown, Martin, Haller & McClain  
; STREET: 1660 Union Street  
; CITY: San Diego  
; STATE: CA  
; COUNTRY: U.S.A.  
; ZIP: 92101-2926  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: Patent In Release #1.0, Version #1.25  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/496,855A  
; FILING DATE: 20-JUN-1995  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: 08-NOV-1993  
; FILING DATE: 08-NOV-1993  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 08/028,031  
; FILING DATE: 08-MAR-1993  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Seidman, Stephanie  
; REGISTRATION NUMBER: 33,779  
; REFERENCE/DOCKET NUMBER: 6362-9369B  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: 619-238-0999  
; TELEFAX: 619-238-0062  
; INFORMATION FOR SEQ ID NO: 1:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 2277 base pairs  
; TYPE: nucleic acid  
; STRANDEDNESS: both  
; TOPOLOGY: both  
; MOLECULE TYPE: cDNA  
; FEATURE:  
; NAME/KEY: CDS  
; LOCATION: 166..1755  
US-08-496-855A-1

Query Match 74.0%; Score 14.8; DB 1; Length 2277;  
Best Local Similarity 88.9%; Pred. No. 1.5e+02;  
Matches 16; Conservative 0; Mismatches 2; Indels 0; Gaps 0;  
QY 3 GGACAGGTAGGGCTTTG 20  
DB 144 GGTCAAGTCAGGGCTTTG 127

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: February 19, 2004, 23:23:34 ; Search time 48.0735 Seconds  
(without alignments)  
1456.787 Million cell updates/sec

Title: US-09-939-853A-140  
Perfect score: 20  
Sequence: 1 ctggacaggttagggctttg 20

Scoring table: IDENTITY\_NUC  
Gapop 10.0 , Gapext 1.0

Searched: 2308684 seqs, 1750822206 residues

Total number of hits satisfying chosen parameters: 4617369

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : Published Applications NA:\*

- 1: /cgn2\_6/ptodata/2/pubpna/US07\_PUBCOMB.seq:\*
- 2: /cgn2\_6/ptodata/2/pubpna/PCT\_NEW\_PUB.seq:\*
- 3: /cgn2\_6/ptodata/2/pubpna/US06\_NEW\_PUB.seq:\*
- 4: /cgn2\_6/ptodata/2/pubpna/US06\_PUBCOMB.seq:\*
- 5: /cgn2\_6/ptodata/2/pubpna/US07\_NEW\_PUB.seq:\*
- 6: /cgn2\_6/ptodata/2/pubpna/PCTUS\_PUBCOMB.seq:\*
- 7: /cgn2\_6/ptodata/2/pubpna/US08\_NEW\_PUB.seq:\*
- 8: /cgn2\_6/ptodata/2/pubpna/US08\_PUBCOMB.seq:\*
- 9: /cgn2\_6/ptodata/2/pubpna/US09A\_PUBCOMB.seq:\*
- 10: /cgn2\_6/ptodata/2/pubpna/US09B\_PUBCOMB.seq:\*
- 11: /cgn2\_6/ptodata/2/pubpna/US09C\_PUBCOMB.seq:\*
- 12: /cgn2\_6/ptodata/2/pubpna/US09\_NEW\_PUB.seq:\*
- 13: /cgn2\_6/ptodata/2/pubpna/US10A\_PUBCOMB.seq:\*
- 14: /cgn2\_6/ptodata/2/pubpna/US10B\_PUBCOMB.seq:\*
- 15: /cgn2\_6/ptodata/2/pubpna/US10C\_PUBCOMB.seq:\*
- 16: /cgn2\_6/ptodata/2/pubpna/US10\_NEW\_PUB.seq:\*
- 17: /cgn2\_6/ptodata/2/pubpna/US60\_NEW\_PUB.seq:\*
- 18: /cgn2\_6/ptodata/2/pubpna/US60\_PUBCOMB.seq:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
C 1	20	100.0	444	9	US-09-867-550-951 Sequence 951, App
C 2	20	100.0	763	9	US-09-867-550-953 Sequence 953, App
C 3	20	100.0	864	10	US-09-814-353-21302 Sequence 21302, A
C 4	17.4	87.0	422	15	US-10-242-535A-25371 Sequence 25371, A
5	16.8	84.0	665	15	US-10-027-632-133814 Sequence 133814, A
C 6	16.8	84.0	2305	15	US-10-094-749-795 Sequence 795, App
C 7	16.8	84.0	2424	15	US-10-027-632-103042 Sequence 103042, A
C 8	16.8	84.0	2424	15	US-10-027-632-103043 Sequence 103043, A
C 9	16.8	84.0	3559	15	US-10-108-260A-602 Sequence 602, App
10	15.8	79.0	403	10	US-09-918-995-35904 Sequence 35904, A
11	15.8	79.0	467	10	US-09-918-995-27228 Sequence 27228, A
12	15.8	79.0	570	15	US-10-027-632-137211 Sequence 137211, A
13	15.8	79.0	663	15	US-10-027-632-208024 Sequence 208024, A
14	15.8	79.0	1152	10	US-09-882-227-2395 Sequence 227, App
15	15.8	79.0	1744	12	US-10-424-599-121358 Sequence 121358, A

16	15.8	79.0	1824	12	US-10-424-599-73670 Sequence 73670, A
17	15.8	79.0	4170	10	US-09-919-039-221 Sequence 221, App
18	15.8	79.0	4170	14	US-10-168-425-23 Sequence 23, Appl
19	15.8	79.0	4359	15	US-10-191-803-341 Sequence 341, App
20	15.8	79.0	24023	13	US-10-094-679-1 Sequence 1, Appli
21	15.8	79.0	99916	9	US-09-816-095-3 Sequence 3, Appli
22	15.4	77.0	673	9	US-09-917-800A-1299 Sequence 1299, Ap
23	15.4	77.0	761	15	US-10-027-632-144930 Sequence 144930, A
24	15.4	77.0	786	15	US-10-027-632-169995 Sequence 169995, A
25	15.4	77.0	786	15	US-10-027-632-169996 Sequence 169996, A
C 26	15.4	77.0	978	15	US-10-027-632-121171 Sequence 121171, A
C 27	15.4	77.0	978	15	US-10-027-632-121172 Sequence 121172, A
28	15.4	77.0	2067	14	US-10-101-510-405 Sequence 405, App
29	15.4	77.0	3685	14	US-10-025-567A-610 Sequence 610, App
C 30	15.4	77.0	4112	10	US-09-930-213-301 Sequence 301, App
31	15.4	77.0	136726	15	US-10-085-117-244 Sequence 244, App
C 32	15.2	76.0	309	9	US-09-836-607-7 Sequence 7, Appli
C 33	15.2	76.0	309	10	US-09-421-112-7 Sequence 7, Appli
C 34	15.2	76.0	320	14	US-10-106-698-3422 Sequence 3422, Ap
C 35	15.2	76.0	430	9	US-09-764-869-408 Sequence 408, App
C 36	15.2	76.0	430	14	US-10-091-504-408 Sequence 408, App
C 37	15.2	76.0	430	15	US-10-027-577-408 Sequence 408, App
C 38	15.2	76.0	468	9	US-09-836-607-34 Sequence 34, Appl
C 39	15.2	76.0	468	10	US-09-421-112-34 Sequence 34, Appl
C 40	15.2	76.0	472	10	US-09-918-995-20214 Sequence 20214, A
C 41	15.2	76.0	475	10	US-09-918-995-1752 Sequence 1752, Ap
C 42	15.2	76.0	478	10	US-09-918-995-32374 Sequence 32374, A
C 43	15.2	76.0	547	15	US-10-027-632-36047 Sequence 36047, A
C 44	15.2	76.0	547	15	US-10-027-632-61127 Sequence 61127, A
C 45	15.2	76.0	547	15	US-10-027-632-303041 Sequence 303041, A

ALIGNMENTS

RESULT 1  
US-09-867-550-951/c  
; Sequence 951, Application US/09867550  
; Patent No. US20020082206A1  
; GENERAL INFORMATION:  
; APPLICANT: Leach, Martin D.  
; APPLICANT: Mehraban, Fuad,  
; APPLICANT: Conley, Pamela  
; APPLICANT: Law, Debbie  
; APPLICANT: Topper, James  
; TITLE OF INVENTION: No. US20020082206A1 Polynucleotides from Atherogenic Cells and F  
; FILE REFERENCE: 21402-013 (Cura-313)  
; CURRENT APPLICATION NUMBER: US/09/867,550  
; CURRENT FILING DATE: 2001-09-20  
; PRIOR APPLICATION NUMBER: US9N 60/208,427  
; PRIOR FILING DATE: 2000-05-30  
; NUMBER OF SEQ ID NOS: 2125  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 951  
; LENGTH: 444  
; TYPE: DNA  
; ORGANISM: Homo sapiens  
US-09-867-550-951

Query Match 100.0%; Score 20; DB 9; Length 444;  
Best Local Similarity 100.0%; Pred. No. 1.1;  
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 CTGACACAGGTAGGGCTTTG 20  
|||||  
DB 60 CTGACACAGGTAGGGCTTTG 41

RESULT 2  
US-09-867-550-953/c  
; Sequence 953, Application US/09867550  
; Patent No. US20020082206A1





; PRIOR APPLICATION NUMBER: US 60/156,358  
; PRIOR FILING DATE: 1999-09-28  
; PRIOR APPLICATION NUMBER: US 60/146,002  
; PRIOR FILING DATE: 1999-08-09  
; NUMBER OF SEQ ID NOS: 325720  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 133814  
; LENGTH: 665  
; TYPE: DNA  
; ORGANISM: Human  
US-10-027-632-133814

Query Match 84.0%; Score 16.8; DB 15; Length 665;  
Best Local Similarity 90.0%; Pred. No. 51;  
Matches 18; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 1 CTGCACAGGTTAGGGCTTTG 20  
|||||  
Db 344 CTGCACAGGATAGGCTGTG 363

RESULT 6  
US-10-094-749-795/c  
; Sequence 795, Application US/10094749  
; Publication No. US20030219741A1  
; GENERAL INFORMATION:  
; APPLICANT: ISOGAI, TAKAO  
; APPLICANT: SUGIYAMA, TOMOYASU  
; APPLICANT: OTSUKI, TETSUJI  
; APPLICANT: WAKAMATSU, AI  
; APPLICANT: SATO, HIROYUKI  
; APPLICANT: ISHII, SHIZUKO  
; APPLICANT: YAMAMOTO, JUN-ICHI  
; APPLICANT: ISONO, YUUKO  
; APPLICANT: HIO, YURI  
; APPLICANT: OTSUKA, KAORU  
; APPLICANT: NAGAI, KEIICHI  
; APPLICANT: IRIE, RYOTARO  
; APPLICANT: TAMECHIKA, ICHIRO  
; APPLICANT: SEKI, NACHIKO  
; APPLICANT: YOSHUKAWA, TSUTOMU  
; APPLICANT: OTSUKA, MOTOTYUKI  
; APPLICANT: NAGAHARI, KENJI  
; APPLICANT: MASUHO, YASUHIKO  
; TITLE OF INVENTION: NOVEL FULL-LENGTH CDNA  
; FILE REFERENCE: 084335/0160  
; CURRENT APPLICATION NUMBER: US/10/094,749  
; CURRENT FILING DATE: 2002-03-12  
; PRIOR FILING DATE: 2002-01-24  
; PRIOR FILING DATE: 2001-09-14  
; NUMBER OF SEQ ID NOS: 3381  
; SOFTWARE: PatentIn Ver. 2.1  
; SEQ ID NO 795  
; LENGTH: 2305  
; TYPE: DNA  
; ORGANISM: Homo sapiens  
US-10-094-749-795

Query Match 84.0%; Score 16.8; DB 15; Length 2305;  
Best Local Similarity 90.0%; Pred. No. 53;  
Matches 18; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 1 CTGCACAGGTTAGGGCTTTG 20  
|||||  
Db 2118 CTGCACAGTTAGGGCTGTG 2099

RESULT 7  
US-10-027-632-103042/c  
; Sequence 103042, Application US/10027632  
; Publication No. US20030204075A9

; GENERAL INFORMATION:  
; APPLICANT: Wang, David G.  
; TITLE OF INVENTION: Identification and Mapping of Single Nucleotide  
; FILE REFERENCE: 108827.129  
; CURRENT APPLICATION NUMBER: US/10/027,632  
; CURRENT FILING DATE: 2002-04-30  
; PRIOR APPLICATION NUMBER: US 60/218,006  
; PRIOR FILING DATE: 2000-07-12  
; PRIOR APPLICATION NUMBER: US 60/198,676  
; PRIOR FILING DATE: 2000-04-20  
; PRIOR APPLICATION NUMBER: US 60/193,483  
; PRIOR FILING DATE: 2000-03-29  
; PRIOR APPLICATION NUMBER: US 60/185,218  
; PRIOR FILING DATE: 2000-02-24  
; PRIOR APPLICATION NUMBER: US 60/167,363  
; PRIOR FILING DATE: 1999-11-23  
; PRIOR APPLICATION NUMBER: US 60/156,358  
; PRIOR FILING DATE: 1999-09-28  
; PRIOR APPLICATION NUMBER: US 60/146,002  
; PRIOR FILING DATE: 1999-08-09  
; NUMBER OF SEQ ID NOS: 325720  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 103042  
; LENGTH: 2424  
; TYPE: DNA  
; ORGANISM: Human  
US-10-027-632-103042

Query Match 84.0%; Score 16.8; DB 15; Length 2424;  
Best Local Similarity 90.0%; Pred. No. 53;  
Matches 18; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 1 CTGCACAGGTTAGGGCTTTG 20  
|||||  
Db 1453 CTGCACAGTTAGGGCTGTG 1434

RESULT 8  
US-10-027-632-103043/c  
; Sequence 103043, Application US/10027632  
; Publication No. US20030204075A9  
; GENERAL INFORMATION:  
; APPLICANT: Wang, David G.  
; TITLE OF INVENTION: Identification and Mapping of Single Nucleotide  
; FILE REFERENCE: 108827.129  
; CURRENT APPLICATION NUMBER: US/10/027,632  
; CURRENT FILING DATE: 2002-04-30  
; PRIOR APPLICATION NUMBER: US 60/218,006  
; PRIOR FILING DATE: 2000-07-12  
; PRIOR APPLICATION NUMBER: US 60/198,676  
; PRIOR FILING DATE: 2000-04-20  
; PRIOR APPLICATION NUMBER: US 60/193,483  
; PRIOR FILING DATE: 2000-03-29  
; PRIOR APPLICATION NUMBER: US 60/185,218  
; PRIOR FILING DATE: 2000-02-24  
; PRIOR APPLICATION NUMBER: US 60/167,363  
; PRIOR FILING DATE: 1999-11-23  
; PRIOR APPLICATION NUMBER: US 60/156,358  
; PRIOR FILING DATE: 1999-09-28  
; PRIOR APPLICATION NUMBER: US 60/146,002  
; NUMBER OF SEQ ID NOS: 325720  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 103043  
; LENGTH: 2424  
; TYPE: DNA  
; ORGANISM: Human  
US-10-027-632-103043

Query Match 84.0%; Score 16.8; DB 15; Length 2424;  
Best Local Similarity 90.0%; Pred. No. 53;

Matches 18; Conservative 0; Mismatches 2; Indels 2; Gaps 0;

QY 1 CTGACACAGTTAGGCTTTG 20  
Db 1453 CTGACACAGTTAGGCTGTG 1434

## RESULT 9

US-10-108-260A-602/c  
; Sequence 602, Application US/10108260A  
; Publication No. US20040005560A1  
; GENERAL INFORMATION:  
; APPLICANT: HELIX RESEARCH INSTITUTE  
; TITLE OF INVENTION: NO. US20040005560A1e1 full length cDNA  
; FILE REFERENCE: HI-A0106  
; CURRENT APPLICATION NUMBER: US/10/108,260A  
; CURRENT FILING DATE: 2002-03-27  
; NUMBER OF SEQ ID NOS: 5458  
; SOFTWARE: PatentIn Ver. 2.1  
; SEQ ID NO 602  
; LENGTH: 3559  
; TYPE: DNA  
; ORGANISM: Homo sapiens  
US-10-108-260A-602

Query Match 84.0%; Score 16.8; DB 15; Length 3559;  
Best Local Similarity 90.0%; Pred. No. 54;  
Matches 18; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 CTGACACAGTTAGGCTTTG 20  
Db 32 CTGACACAGTTAGGCTTTG 13

## RESULT 10

US-09-918-995-35904  
; Sequence 35904, Application US/09918995  
; Publication No. US20030073623A1  
; GENERAL INFORMATION:  
; APPLICANT: Hyseq, Inc.  
; TITLE OF INVENTION: NOVEL NUCLEIC ACID SEQUENCES OBTAINED  
; FILE REFERENCE: 20411-756  
; CURRENT APPLICATION NUMBER: US/09/918,995  
; CURRENT FILING DATE: 2001-07-30  
; PRIOR APPLICATION NUMBER: US/09/235,076  
; PRIOR FILING DATE: 1999-01-20  
; NUMBER OF SEQ ID NOS: 38054  
; SOFTWARE: FastSeq for Windows Version 3.0  
; SEQ ID NO 35904  
; LENGTH: 403  
; TYPE: DNA  
; ORGANISM: Homo sapiens  
; NAME/KEY: misc\_feature  
; LOCATION: (1)...(403)  
; OTHER INFORMATION: n = A,T,C or G  
US-09-918-995-35904

Query Match 79.0%; Score 15.8; DB 10; Length 403;  
Best Local Similarity 89.5%; Pred. No. 1.7e+02;  
Matches 17; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 CTGACACAGTTAGGCTTT 19  
Db 48 CTGACACAGTTAGGCTTT 66

## RESULT 11

US-09-918-995-27228  
; Sequence 27228, Application US/09918995  
; Publication No. US20030073623A1  
; GENERAL INFORMATION:

; APPLICANT: Hyseq, Inc.  
; TITLE OF INVENTION: NOVEL NUCLEIC ACID SEQUENCES OBTAINED  
; FILE REFERENCE: 20411-756  
; CURRENT APPLICATION NUMBER: US/09/918,995  
; CURRENT FILING DATE: 2001-07-30  
; PRIOR APPLICATION NUMBER: US/09/235,076  
; PRIOR FILING DATE: 1999-01-20  
; NUMBER OF SEQ ID NOS: 38054  
; SOFTWARE: FastSeq for Windows Version 3.0  
; SEQ ID NO 27228  
; LENGTH: 467  
; TYPE: DNA  
; ORGANISM: Homo sapiens  
; NAME/KEY: misc\_feature  
; LOCATION: (1)...(467)  
; OTHER INFORMATION: n = A,T,C or G  
US-09-918-995-27228

Query Match 79.0%; Score 15.8; DB 10; Length 467;  
Best Local Similarity 89.5%; Pred. No. 1.7e+02;  
Matches 17; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 CTGACACAGTTAGGCTTT 19  
Db 152 CTGACACAGTTAGGCTTT 170

## RESULT 12

US-10-027-632-137211  
; Sequence 137211, Application US/10027632  
; Publication No. US20030204075A9  
; GENERAL INFORMATION:  
; APPLICANT: Wang, David G.  
; TITLE OF INVENTION: Identification and Mapping of Single Nucleotide  
; FILE REFERENCE: 108827.129  
; CURRENT APPLICATION NUMBER: US/10/027,632  
; CURRENT FILING DATE: 2002-04-30  
; PRIOR APPLICATION NUMBER: US 60/218,006  
; PRIOR FILING DATE: 2000-07-12  
; PRIOR APPLICATION NUMBER: US 60/198,676  
; PRIOR FILING DATE: 2000-04-20  
; PRIOR APPLICATION NUMBER: US 60/193,483  
; PRIOR FILING DATE: 2000-03-29  
; PRIOR APPLICATION NUMBER: US 60/185,218  
; PRIOR FILING DATE: 2000-02-24  
; PRIOR APPLICATION NUMBER: US 60/167,363  
; PRIOR FILING DATE: 1999-11-23  
; PRIOR APPLICATION NUMBER: US 60/156,358  
; PRIOR FILING DATE: 1999-09-28  
; PRIOR APPLICATION NUMBER: US 60/146,002  
; NUMBER OF SEQ ID NOS: 325720  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 137211  
; LENGTH: 570  
; TYPE: DNA  
; ORGANISM: Human  
US-10-027-632-137211

Query Match 79.0%; Score 15.8; DB 15; Length 570;  
Best Local Similarity 89.5%; Pred. No. 1.7e+02;  
Matches 17; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 CTGACACAGTTAGGCTTT 19  
Db 301 CAGGACAGATTAGGCTTT 319

## RESULT 13

US-10-027-632-208024



GenCore version 5.1.6  
Copyright (c) 1993 - 2004 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: February 19, 2004, 21:21:18 ; Search time 113.831 Seconds  
(without alignments)  
9899.970 Million cell updates/sec

Title: US-09-939-853a-141

Perfect score: 26  
Sequence: 1 cttcttggaagtctgccagtgccctt 26

Scoring table: IDENTITY NUC  
Gapop 10.0 , Gapext 1.0

Searched: 3470272 seqs, 21671516995 residues

Total number of hits satisfying chosen parameters: 6940544

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : GenEmbl.\*

- 1: gb.ba.\*
- 2: gb.htg.\*
- 3: gb.in.\*
- 4: gb.om.\*
- 5: gb.ov.\*
- 6: gb.pat.\*
- 7: gb.ph.\*
- 8: gb.pl.\*
- 9: gb.pr.\*
- 10: gb.ro.\*
- 11: gb.sts.\*
- 12: gb.sy.\*
- 13: gb.un.\*
- 14: gb.vi.\*
- 15: em.ba.\*
- 16: em.fun.\*
- 17: em.hum.\*
- 18: em.in.\*
- 19: em.mu.\*
- 20: em.om.\*
- 21: em.or.\*
- 22: em.ov.\*
- 23: em.pat.\*
- 24: em.ph.\*
- 25: em.pl.\*
- 26: em.ro.\*
- 27: em.sts.\*
- 28: em.un.\*
- 29: em.vi.\*
- 30: em.htg.hum.\*
- 31: em.htg.in.\*
- 32: em.htg.other.\*
- 33: em.htg.mus.\*
- 34: em.htg.pln.\*
- 35: em.htg.rod.\*
- 36: em.htg.mam.\*
- 37: em.htg.vrt.\*
- 38: em.sy.\*
- 39: em.htgo.hum.\*
- 40: em.htgo.mus.\*
- 41: em.htgo.other.\*

Pred. No. is the number of results predicted by chance to have a

score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Match	Length	DB	ID	Description
1	26	100.0	26	6	AX443200	AX443200 Sequence
2	26	100.0	1183	6	AX443133	AX443133 Sequence
3	26	100.0	1183	6	AX443135	AX443135 Sequence
4	26	100.0	2538	9	BC042041	BC042041 Homo sapi
5	26	100.0	2567	6	AX452880	AX452880 Sequence
6	26	100.0	2788	6	AX780857	AX780857 Sequence
7	26	100.0	66741	9	HS460J8	AL031662 Human DNA
8	26	100.0	145833	2	AC026539	AC026539 Homo sapi
9	21.2	81.5	177004	2	AC129114	AC129114 Rattus no
10	21.2	81.5	179497	9	AC020636	AC020636 Homo sapi
11	21.2	81.5	259967	2	AC105625	AC105625 Rattus no
12	21.2	81.5	261492	2	AC095338	AC095338 Rattus no
13	20.8	80.0	76332	2	AC021312	AC021312 Homo sapi
14	20.8	80.0	184541	9	AC011405	AC011405 Homo sapi
15	20.4	78.5	120194	9	AL355474	AL355474 Human DNA
16	20.4	78.5	157289	2	AC021447	AC021447 Homo sapi
17	20.4	78.5	183132	9	AC025038	AC025038 Homo sapi
18	20.2	77.7	1617	8	AK071089	AK071089 Oryza sat
19	20.2	77.7	2808	8	AK072171	AK072171 Oryza sat
20	20.2	77.7	110000	2	AC106661.2	Continuation (3 of
21	20.2	77.7	110000	2	AC127890.2	Continuation (3 of
22	20.2	77.7	130616	2	AP003950	AP003950 Oryza sat
23	20.2	77.7	139848	2	AP004335	AP004335 Oryza sat
24	20.2	77.7	148542	8	AP005127	AP005127 Oryza sat
25	20.2	77.7	163096	9	CNS01D01	AL133167 Human chr
26	20.2	77.7	166305	9	AC012447	AC012447 Homo sapi
27	20.2	77.7	167551	2	AC112819	AC112819 Rattus no
28	20.2	77.7	168135	2	AC120654	AC120654 Rattus no
29	20.2	77.7	225443	2	AC137194	AC137194 Rattus no
30	20.2	77.7	246040	2	AC103184	AC103184 Rattus no
31	20.2	77.7	283262	2	AC095135	AC095135 Rattus no
32	19.8	76.2	32587	9	AC096665	AC096665 Homo sapi
33	19.8	76.2	165349	2	AC093665	AC093665 Homo sapi
34	19.8	76.2	170240	9	AL160279	AL160279 Human DNA
35	19.8	76.2	194998	10	AL627075	AL627075 Mouse DNA
36	19.8	76.2	210164	2	AC010732	AC010732 Homo sapi
37	19.8	76.2	241802	2	AC147280	AC147280 Pan trogl
38	19.6	75.4	6159	4	SSC133742	AL133742 Sus scrof
39	19.6	75.4	139958	2	AC141936	AC141936 Rattus no
40	19.6	75.4	174987	2	AC145778	AC145778 Sus scrof
41	19.6	75.4	183097	9	AC087358	AC087358 Homo sapi
42	19.6	75.4	183360	2	AC022525	AC022525 Homo sapi
43	19.6	75.4	184160	2	AC147367	AC147367 Mus muscu
44	19.6	75.4	209380	10	AC133186	AC133186 Mus muscu
45	19.6	75.4	255608	2	AC109405	AC109405 Rattus no

## ALIGNMENTS

RESULT 1	AX443200	Sequence 141 from Patent WO0216599.	26 bp	DNA	linear	PAT 02-JUL-2002
LOCUS	AX443200	Sequence 141 from Patent WO0216599.				
DEFINITION	AX443200	Sequence 141 from Patent WO0216599.				
ACCESSION	AX443200	Sequence 141 from Patent WO0216599.				
VERSION	AX443200.1	GI:21690595				
KEYWORDS		synthetic construct				
SOURCE		synthetic construct				
ORGANISM		artificial sequences.				
REFERENCE		1				
AUTHORS		Burgess,C.E., Conley,P.B., Grosse,W.M., Hart,M., Kekuda,R., Shimkets,R.A., Spytek,K.A., Szekeres,E.S., Tomlinson,J.E., Topper,J.N. and Yang,R.B.				
TITLE		Proteins and nucleic acids encoding same				

JOURNAL Patent: WO 0216599-A 141 28-FEB-2002;  
Curagen Corporation (US); COR THERAPEUTICS, INC. (US)  
FEATURES Location/Qualifiers  
source

1..26  
/organism="synthetic construct"  
/mol\_type="unassigned DNA"  
/db\_xref="taxon:32630"  
/note="oligonucleotide primer"

## ORIGIN

Query Match 100.0%; Score 26; DB 6; Length 26;  
Best Local Similarity 100.0%; Pred. No. 0.56; Indels 0; Gaps 0;  
Matches 26; Conservative 0; Mismatches 0;

Qy 1 CCTTCTGGAAGTCTGCCAGTGTCTT 26  
|||||  
Db 1 CCTTCTGGAAGTCTGCCAGTGTCTT 26

RESULT 2  
AX443133/c 1183 bp DNA linear PAT 02-JUL-2002  
LOCUS  
DEFINITION Sequence 74 from Patent WO0216599.  
ACCESSION AX443133  
VERSION AX443133.1 GI:21690555  
KEYWORDS  
SOURCE Homo sapiens (human)

ORGANISM  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
REFERENCE  
AUTHORS Burgess, C.E., Conley, P.B., Grosse, W.M., Hart, M., Kekuda, R.,  
Shimkets, R.A., Spytek, K.A., Szekeres, E.S., Tomlinson, J.E.,  
Topper, J.N. and Yang, R.B.

TITLE  
Proteins and nucleic acids encoding same

JOURNAL  
Patent: WO 0216599-A 74 28-FEB-2002;  
Curagen Corporation (US); COR THERAPEUTICS, INC. (US)  
FEATURES Location/Qualifiers

1..1183  
/organism="Homo sapiens"  
/mol\_type="unassigned DNA"  
/db\_xref="taxon:9606"

## ORIGIN

Query Match 100.0%; Score 26; DB 6; Length 1183;  
Best Local Similarity 100.0%; Pred. No. 0.37; Indels 0; Gaps 0;  
Matches 26; Conservative 0; Mismatches 0;

Qy 1 CCTTCTGGAAGTCTGCCAGTGTCTT 26  
|||||  
Db 276 CCTTCTGGAAGTCTGCCAGTGTCTT 251

RESULT 3  
AX443135 1183 bp DNA linear PAT 02-JUL-2002  
LOCUS  
DEFINITION Sequence 76 from Patent WO0216599.  
ACCESSION AX443135  
VERSION AX443135.1 GI:21690556  
KEYWORDS  
SOURCE Homo sapiens (human)

ORGANISM

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
REFERENCE  
AUTHORS Burgess, C.E., Conley, P.B., Grosse, W.M., Hart, M., Kekuda, R.,  
Shimkets, R.A., Spytek, K.A., Szekeres, E.S., Tomlinson, J.E.,  
Topper, J.N. and Yang, R.B.

TITLE  
Proteins and nucleic acids encoding same

JOURNAL  
Patent: WO 0216599-A 76 28-FEB-2002;  
Curagen Corporation (US); COR THERAPEUTICS, INC. (US)  
FEATURES Location/Qualifiers

1..1183

/organism="Homo sapiens"  
/mol\_type="unassigned DNA"  
/db\_xref="taxon:9606"

## ORIGIN

Query Match 100.0%; Score 26; DB 6; Length 1183;  
Best Local Similarity 100.0%; Pred. No. 0.37; Indels 0; Gaps 0;  
Matches 26; Conservative 0; Mismatches 0;

Qy 1 CCTTCTGGAAGTCTGCCAGTGTCTT 26  
|||||  
Db 908 CCTTCTGGAAGTCTGCCAGTGTCTT 933

RESULT 4  
BC042041/c 2538 bp mRNA linear PRI 07-OCT-2003  
LOCUS  
DEFINITION Homo sapiens Src-like-adaptor 2, transcript variant 1, mRNA (cdna  
clone MGC:49845 IMAGE:4429896), complete cds.  
ACCESSION BC042041  
VERSION BC042041.1 GI:27469842  
KEYWORDS MGC.  
SOURCE Homo sapiens (human)

ORGANISM  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
REFERENCE  
AUTHORS Strausberg, R.L., Feingold, E.A., Grouse, L.H., Derge, J.G.,  
Klausner, R.D., Collins, P.S., Wagner, L., Shenman, C.M., Schuler, G.D.,  
Altschul, S.F., Zeeberg, B., Buetow, K.H., Schaefer, C.F., Bhat, N.K.,  
Hopkins, R.F., Jordan, H., Moore, T., Max, S.I., Wang, J., Haieh, F.,  
Diatchenko, L., Marusina, K., Farmer, A.A., Rubin, G.M., Hong, L.,  
Stapleton, M., Soares, M.B., Bonaldo, M.F., Casavant, T.L.,  
Scheetz, T.E., Brownstein, M.J., Usdin, T.B., Toshiyuki, S.,  
Carninci, P., Prange, C., Raha, S.S., Loquellano, N.A., Peters, G.J.,  
Abramson, R.D., Mullahy, S.J., Bosak, S.A., McEwan, P.J.,  
McKernan, K.J., Malek, J.A., Gunaratne, P.H., Richards, S.,  
Worley, K.C., Hale, S., Garcia, A.M., Gay, L.J., Hulyk, S.W.,  
Villalón, D.K., Muzny, D.M., Sodergren, E.J., Lu, X., Gibbs, R.A.,  
Fahey, J., Helton, E., Kettman, M., Madan, A., Rodriguez, S.,  
Sanchez, A., Whitting, M., Madan, A., Young, A.C., Shevchenko, Y.,  
Bouffard, G.G., Blakesley, R.W., Touchman, J.W., Green, E.D.,  
Dickson, M.C., Rodriguez, A.C., Grimwood, J., Schmutz, J., Myers, R.M.,  
Butterfield, Y.S., Krzywicki, M.I., Skalska, U., Smal, D.E.,  
Schnerch, A., Schein, J.E., Jones, S.J. and Marra, M.A.

Generation and initial analysis of more than 15,000 full-length  
human and mouse cDNA sequences  
Proc. Natl. Acad. Sci. U.S.A. 99 (26), 16899-16903 (2002)  
12477932  
2 (bases 1 to 2538)  
Strausberg, R.  
Direct Submission  
Submitted (23-DEC-2002) National Institutes of Health, Mammalian  
Gene Collection (MGC), Cancer Genomics Office, National Cancer  
Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590,  
USA  
NIH-MGC Project URL: <http://mgc.nci.nih.gov>  
Contact: MGC help desk  
Email: [cgapbs-remail.nih.gov](mailto:cgapbs-remail.nih.gov)  
Tissue Procurement: DCID/DTP  
cDNA Library Preparation: Life Technologies, Inc.  
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (ILNL)  
DNA Sequencing by: Sequencing Group at the Stanford Human Genome  
Center, Stanford University School of Medicine, Stanford, CA 94305  
Web site: <http://www-shgc.stanford.edu>  
Contact: (Dickson, Mark) [mdc@paxil.stanford.edu](mailto:mdc@paxil.stanford.edu)  
Dickson, M., Schmutz, J., Grimwood, J., Rodriguez, A., and Myers,  
R. M.

TITLE  
JOURNAL  
MEDLINE  
PUBMED  
REFERENCE  
AUTHORS  
TITLE  
JOURNAL

REMARK  
COMMENT

Clone distribution: MGC clone distribution information can be found  
through the I.M.A.G.E. Consortium/ILNL at: <http://image.llnl.gov>  
Series: IRAC Plate: 88 Row: a Column: 20

This clone was selected for full length sequencing because it passed the following selection criteria: matched mRNA gi: 28416422.

```

FEATURES
    source
        1..2538
            /organism="Homo sapiens"
            /mol_type="mRNA"
            /db_xref="taxon:9606"
            /clone="MGC:49845 IMAGE:4429896"
            /tissue_type="Prostate, adenocarcinoma."
            /clone_lib="NIH_MGC_91"
            /lab_host="DH10B"
            /note="Vector: pCMV-SPORT6"
    gene
        1..2538
            /gene="SLA2"
            /note="synonyms: FLJ21992, SLAP-2, C20orf156, MGC49845, SLAP2"
            /db_xref="LocusID:84174"
            /db_xref="MIM:606577"
    CDS
        363..1148
            /codon_start=1
            /product="Src-like-adaptor 2, isoform a"
            /protein_id="AA42041.1"
            /db_xref="GI:27469843"
            /db_xref="LocusID:84174"
            /translation="MGSLSRRKSLPSLSVSSVQGGPVTMEASRKATAVALGSPF
            AGPALSRLGLGELTIVSDGWTVLSEVSGREYNIPSVHAKVSHGWLVEGLSRE
            KABELLLPQNGEPLTIVSDGWTVLSEVSGREYNIPSVHAKVSHGWLVEGLSRE
            SPLTPPSQALVDHSELADDCILKSPCVLQRPAGPLPGKIDPLPTVQRTPLNWK
            ELDSLLFSEAAATGERSLSEGLRESLSYIISLNDFAVSLDDA"
            /misc_feature
                469..632
                    /notes="SH3; Region: SH3 domain. SH3 (Src homology 3)
                    domains are often indicative of a protein involved in
                    signal transduction related to cytoskeletal organization.
                    First described in the Src cytoplasmic tyrosine kinase.
                    The structure is a partly opened beta barrel"
                    /db_xref="CDD:pfam00018"
            /misc_feature
                642..890
                    /notes="SH2; Region: SH2 domain"
                    /db_xref="CDD:pfam00017"
    ORIGIN
        Query Match          100.0%; Score 26; DB 9; Length 2538;
        Best Local Similarity 100.0%; Pred. No. 0.34; Indels 0; Gaps 0;
        Matches 26; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

        QY 1 CCTCTGGAAGTCTGCCAGTGTCTT 26
            |||||
        Db 241 CCTCTGGAAGTCTGCCAGTGTCTT 216
            |||||

    RESULT 5
    AX452880/c
    LOCUS
    DEFINITION
        Human DNA sequence from clone RP3-460J8 on chromosome
        20q11.21-11.23 Contains the 3' end of the gene for a novel protein
        similar to N-myc downstream regulated (NDRG1) the 5' end of a gene
        encoding a novel protein tyrosine kinase, ESTs, STSS and GSSs,
        complete sequence.
    ACCESSION
        AL031662
    VERSION
        AL031662.26 GI:9716901
    KEYWORDS
        HTG; NDRG1; SH2 domain.
    SOURCE
        Homo sapiens (human)
    ORGANISM
        Homo sapiens
    Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
    Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
    Chang, H., Yang, W.P., Wu, Y., Whitney, G.S., Perez-Villar, J.J. and
    Kanner, S.B.
    Cloning and expression of human slap-2: a novel sh2/sh3
    domain-containing human slap homologue having immune cell-specific
    expression
    Patent: WO 0242457-A 1 30-MAY-2002;
    Bristol-Myers Squibb Co. (US)
    Location/Qualifiers
        1..2567
            /organism="Homo sapiens"
            /mol_type="unassigned DNA"

    /db_xref="taxon:9606"

    ORIGIN
        Query Match          100.0%; Score 26; DB 6; Length 2788;
        Best Local Similarity 100.0%; Pred. No. 0.33; Indels 0; Gaps 0;
        Matches 26; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

        QY 1 CCTCTGGAAGTCTGCCAGTGTCTT 26
            |||||
        Db 265 CCTCTGGAAGTCTGCCAGTGTCTT 240
            |||||

    RESULT 7
    HS460J8
    LOCUS
    DEFINITION
        Human DNA sequence from clone RP3-460J8 on chromosome
        20q11.21-11.23 Contains the 3' end of the gene for a novel protein
        similar to N-myc downstream regulated (NDRG1) the 5' end of a gene
        encoding a novel protein tyrosine kinase, ESTs, STSS and GSSs,
        complete sequence.
    ACCESSION
        AL031662
    VERSION
        AL031662.26 GI:9716901
    KEYWORDS
        HTG; NDRG1; SH2 domain.
    SOURCE
        Homo sapiens (human)
    ORGANISM
        Homo sapiens
    Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
    Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
    Skuce, C.
    Direct Submission
    Submitted (23-JUL-2001) Sanger Centre, Hinxton, Cambridgeshire,
    CB10 1SA, UK. E-mail enquiries: humquery@sanger.ac.uk
    requests: clonerequest@sanger.ac.uk
    On Aug 7, 2000 this sequence version replaced gi:6425549.
    During sequence assembly data is compared from overlapping clones.
    Where differences are found these are annotated as variations
    together with a note of the overlapping clone name. Note that the
    variation annotation may not be found in the sequence submission

```

corresponding to the overlapping clone, as we submit sequences with only a small overlap as described above.

The following abbreviations are used to associate primary accession numbers given in the feature table with their source databases: Em, EMBL; Sw, SWISSPROT; Tr, TREMBL; Wp, WORMPEP; information on the WORMPEP database can be found at [http://www.sanger.ac.uk/Projects/C\\_elegans/wormpep](http://www.sanger.ac.uk/Projects/C_elegans/wormpep) This sequence was generated from part of bacterial clone contigs of human chromosome 20, constructed by the Sanger Centre Chromosome 20 Mapping Group. Further information can be found at <http://www.sanger.ac.uk/HGP/Chr20>

IMPORTANT: This sequence is not the entire insert of clone RP3-460J8. It may be shorter because we sequence overlapping sections only once, except for a 100 base overlap.

The true right end of clone RP3-460J8 is at 66741 in this sequence. The true left end of clone RP3-469A13 is at 41767 in this sequence. This sequence was finished as follows unless otherwise noted: all regions were either double-stranded or sequenced with an alternate chemistry or covered by high quality data (i.e., phred quality >= 30); an attempt was made to resolve all sequencing problems, such as compressions and repeats; all regions were covered by at least one plasmid subclone or more than one M13 subclone; and the assembly was confirmed by restriction digest. RP3-460J8 is from the library RP3-3 constructed by the group of Pieter de Jong. For further details see <http://www.chori.org/bacpac/home.htm>

VECTOR: pCYPAC2.

## FEATURES

```

source
    Location/Qualifiers
        1..66741
            /organism="Homo sapiens"
            /mol_type="genomic DNA"
            /db_xref="taxon:9606"
            /chromosome="20"
            /map="q11.21-11.23"
            /clone="RP3-460J8"
            /clone_lib="RP3-3"
            complement(50..544)
            /note="match: GSS: Em:B45150"
            /gene="dJ977B1.1"
            complement(240..339,6995..7128)
            /gene="dJ977B1.1"
            complement(dJ977B1.1)
            /product="dJ460J8.2 (novel protein tyrosine kinase with Src homology 2 (SH2) domain)"
            /note="match: cDNAs: Em:AK025645
            match: ESTs: Em:BG178487"
            /evidence=not experimental
            complement(join(<240..339,6995..7085))
            /gene="dJ977B1.1"
            /note="Continues in Em:AL050318 as dJ977B1.2"
            /codon_start=1
            /evidence=not experimental
            /product="dJ460J8.2 (novel protein tyrosine kinase with Src homology 2 (SH2) domain)"
            /protein_id="CAC44645.1"
            /db_xref="GI:15020830"
            /db_xref="GOA:Q9H6Q3"
            /translation="MGSLPGRKSLPSPSSVQGGPVTMEARSKATAVALGSPF
            AGPAELSLRLGEPLTIVS"
            536..565
repeat_region
    /note="15 copies 2 mer ag 93% conserved"
repeat_region
    2375..2402
    /note="14 copies 2 mer ta 100% conserved"
misc_feature
    2783..3155
    /note="match: GSS: Em:AQ0807191"
    complement(6983..7482)
    /note="match: GSS: Em:AQ556467"
    complement(6995..7492)
    /note="match: GSS: Em:AQ556478"
    7486..7995
    /note="match: STS: Em:HS427J18"
repeat_region
    10615..10640
    /note="13 copies 2 mer tt 92% conserved"
repeat_region
    13593..13961
    /note="MER21B repeat: matches 422..785 of consensus"
repeat_region
    14248..14297
    /note="MER21B repeat: matches 374..422 of consensus"
repeat_region
    14594..14614
    /note="MER21B repeat: matches 355..374 of consensus"
repeat_region
    14746..15078
    /note="MER21B repeat: matches 1..355 of consensus"
    complement(17516)
    17519..17886
    /note="match: GSS: Em:A2067993"
    complement(17521)
    complement(17522)
    complement(17523..54534)
    /gene="dJ469A13.3"
    complement(join(<17523..19451,20551..20589,22110..22161,
    22247..22282,26087..26134,30796..30847,30986..31051,
    32039..32142,37091..37147,46566..46652,48268..48328,
    50163..50225,53242..53362,54429..54534))
    /gene="dJ469A13.3"
    /product="dJ460J8.1 (continued from dJ469A13.3 in
    Em:AL132768)"
    /note="match: cDNAs: Em:AB033922 Em:U52073 Em:D87953
    Em:X92845 Em:AF004162 Em:U60593 Em:AF045564 Em:AB033921
    Em:AF159092 Em:AF147402 Em:M59814
    match: ESTs: Em:AA718726 Em:AA039000 Em:AW003952 Em:W99263
    Em:AA325826 Em:A1230982 Em:AA162360 Em:AA445016
    Em:AV002395 Em:T85147 Em:A1786615 Em:AA113437 Em:A1004026
    Em:T88705 Em:A1786673 Em:A1181197 Em:AA764653 Em:AA316771
    Em:A0035165 Em:AV002368"
    /evidence=not experimental
    complement(17542..17547)
    /gene="dJ469A13.3"
    17555..17811
    /note="match: STS: Em:G19945"
    join(19202..19451,22110..22282,26087..26127,30796..30847)
    /note="match: STS: Em:G23762"
    complement(join(19270..19451,20551..20589,22110..22161,
    22247..22282,26087..26134,30796..30847,30986..31051,
    32039..32142,37091..37147,46566..46652,48268..48328,
    50163..50225,53242..53362,54429..54534))
    /gene="dJ469A13.3"
    /note="novel protein (FLJ13556) similar to N-myc
    downstream regulated (NDRG1)
    match: proteins: Sw:Q62433 Sw:Q92597 Tr:Q9Z2L9 Sw:P97862"
    /codon_start=1
    /evidence=not experimental
    /product="dJ460J8.1 (continued from dJ469A13.3 in
    Em:AL132768)"
    /protein_id="CAB65625.1"
    /db_xref="GI:6687781"
    /db_xref="GOA:Q9UGV2"
    /db_xref="SWISS-PROT:Q9UGV2"
    /translation="EHDIETHGVVHTIRGLPKGNRPVLTTHDIGNHXSCNAPF
    NFDNQRTQHFVCHVDAPQCGAFSPFGYQIPTWDELAEMLPVLTTHLSKSLII
    GIVGAGAYILSRFALHAPLPELVGLINVDPCAKGMDWAASKLGLTTNVVDIILA
    HFGQEEQLANLDIQTRMHIAQDINQNLQLEFINSYGRDLEIERPILGQNDKLS
    KTLKCSLLVVGDSPTAVEAVECNSRLNPINTLLXWADCGGLPVQVQGLTEAFK
    YFLOCMGYIPYVQLSHLSTGSPVPSAMTRLSRTHSTSSSLGSGESPFSSRSVTSNQS
    DGTQESCSDPVLDRHQTMVESC"
    complement(24246..24780)
    /gene="dJ469A13.3"
    /note="match: GSS: Em:AQ592789"
    complement(27980..28556)
    /gene="dJ469A13.3"
    /note="match: GSS: Em:AQ308867"
    34313..34420
    /note="HY1 repeat: matches 1..109 of consensus"
    34505..34560
    /note="28 copies 2 mer ta 78% conserved"
    34564..34611

```



```

/note="24 copies 2 mer ta 79% conserved"
42533..43658
/note="MER52C repeat: matches 1..1278 of consensus"
47989..49470
/note="match: GSS: Em:AQ672417"
49391..49480
/note="45 copies 2 mer ta 76% conserved"
52922..53120
/note="match: STS: Em:G04621"
/complement(53426..54032)
/note="match: GSS: Em:AQ314824"
54054..54477
/note="match: GSS: Em:AQ310681"
57509..57838
/note="match: STS: Em:G07504"
/complement(57696..58013)
/note="match: STS: Em:G07632 Em:G07634"
58856..59001
/note="73 copies 2 mer at 83% conserved"
/complement(58930..59281)
/note="match: GSS: Em:AQ067563"
59062..59236
/note="TIGER2 repeat: matches 2541..2714 of consensus"
59738..60495
/note="TIGER2 repeat: matches 1780..2541 of consensus"
60719..61831
/note="MER11C repeat: matches 1..1071 of consensus"
62547..63174

Query Match 100.0%; Score 26; DB 9; Length 66741;
Best Local Similarity 100.0%; Pred. No. 0.23;
Matches 26; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 CCTCTGGAAGTCTGCCAGTGCCTT 26
Db 11665 CCTCTGGAAGTCTGCCAGTGCCTT 11690

RESULT 8
AC026539/c
LOCUS
DEFINITION Homo sapiens chromosome 20 clone RP11-712N14 map 20, WORKING DRAFT
SEQUENCE, 37 unordered pieces.
AC026539 GI:7656813
VERSION HTG; HTGS_PHASE1; HTGS_DRAFT.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
Birren, B., Linton, L., Nusbaum, C. and Lander, E.
1 (bases 1 to 145833)
Homo sapiens chromosome 20, clone RP11-712N14
Unpublished
REFERENCE 2 (bases 1 to 145833)
Birren, B., Linton, L., Nusbaum, C., Lander, E., Abraham, H., Allen, N.,
Anderson, S., Baldwin, J., Barna, N., Bastien, V., Beda, F.,
Boguski, L., Boukagaiter, B., Brown, A., Burkett, G.,
Campopiano, A., Castle, A., Choquel, Y., Colangelo, M., Collins, S.,
Collamore, A., Cooke, P., DeArrellano, K., Dewar, K., Diaz, J.S.,
Dodge, S., Domino, M., Doyle, M., Ferreira, P., FitzHugh, W., Gage, D.,
Galagan, J., Gardyna, S., Ginde, S., Govette, M., Graham, L.,
Grand-Pierre, N., Grant, G., Hagos, B., Heaford, A., Horton, L.,
Howland, J.C., Iliev, I., Johnson, R., Jones, C., Kann, L., Karatas, A.,
Klein, J., Laroque, K., Lamazares, R., Landers, T., Lehoczy, J.,
Levine, R., Lieu, C., Liu, G., Locke, K., Macdonald, P., Marquis, N.,
McCarthy, M., McEwan, P., McGurk, A., McKernan, K., MCPheeters, R.,
Meldrum, J., Meneses, L., Mihova, T., Miranda, C., Mlenga, V., Morrow, J.,
Murphy, T., Naylor, J., Norman, C.H., O'Connor, T., O'Donnell, P.,
O'Neil, D., Oliver, T.M., Oliver, J., Peterson, K., Pierre, N.,
Pisani, C., Pollara, V., Raymond, C., Riley, R., Rogov, P., Rothman, D.,

```

TITLE  
JOURNAL

## COMMENT

Roy, A., Santos, R., Schauer, S., Severy, P., Spencer, B.,  
 Stange-Thomann, N., Stojanovic, N., Subramanian, A., Talamas, J.,  
 Tesfaye, S., Theodore, J., Tirrell, A., Travers, M., Trigilio, J.,  
 Vassiliev, H., Viel, R., Vo, A., Wilson, B., Wu, X., Wyman, D., Ye, W.J.,  
 Young, G., Zainoun, J., Zimmer, A. and Zody, M.  
 Direct Submission  
 Submitted (22-MAR-2000) Whitehead Institute/MIT Center for Genome  
 Research, 320 Charles Street, Cambridge, MA 02141, USA  
 On Apr 27, 2000 this sequence version replaced gi:7283243.  
 All repeats were identified using RepeatMasker:  
 Smit, A.F.A. & Green, P. (1996-1997)  
 http://ftp.genome.washington.edu/RM/RepeatMasker.html

----- Genome Center  
 Center: Whitehead Institute/ MIT Center for Genome Research

Center code: WIBR

Web site: http://www-seq.wi.mit.edu

Contact: sequence.submissions@genome.wi.mit.edu

----- Project Information

Center project name: L7115

Center clone name: 712\_N14

----- Summary Statistics

Sequencing vector: M13; M77815; 100% of reads

Chemistry: Dye-terminator Big Dye; 100% of reads

Assembly program: Phrap; version 0.960731

Consensus quality: 125577 bases at least Q40

Consensus quality: 135703 bases at least Q30

Consensus quality: 135593 bases at least Q20

Insert size: 182000; agarose-fp

Insert size: 142233; sum-of-contigs

Quality coverage: 2.6 in Q20 bases; agarose-fp

Quality coverage: 3.3 in Q20 bases; sum-of-contigs

\* NOTE: This is a 'working draft' sequence. It currently  
 \* consists of 37 contigs. The true order of the pieces  
 \* is not known and their order in this sequence record is  
 \* arbitrary. Gaps between the contigs are represented as  
 \* runs of N, but the exact sizes of the gaps are unknown.  
 \* This record will be updated with the finished sequence  
 \* as soon as it is available and the accession number will  
 \* be preserved.

1 1214: contig of 1214 bp in length  
 \* 1215: gap of 100 bp  
 \* 1315: contig of 1359 bp in length  
 \* 2673: gap of 100 bp  
 \* 2773: gap of 100 bp  
 \* 4520: contig of 1747 bp in length  
 \* 4620: gap of 100 bp  
 \* 4621: contig of 1341 bp in length  
 \* 5961: contig of 1341 bp in length  
 \* 6061: gap of 100 bp  
 \* 6082: contig of 1658 bp in length  
 \* 7719: gap of 100 bp  
 \* 7819: gap of 100 bp  
 \* 7820: contig of 1980 bp in length  
 \* 9899: gap of 100 bp  
 \* 9900: contig of 1535 bp in length  
 \* 11434: contig of 1535 bp in length  
 \* 11534: gap of 100 bp  
 \* 14382: contig of 2848 bp in length  
 \* 14482: gap of 100 bp  
 \* 17148: contig of 2666 bp in length  
 \* 17149: gap of 100 bp  
 \* 17249: contig of 1991 bp in length  
 \* 19239: gap of 100 bp  
 \* 19339: contig of 1763 bp in length  
 \* 21103: contig of 100 bp  
 \* 21202: gap of 100 bp  
 \* 23371: contig of 2169 bp in length  
 \* 23471: gap of 100 bp  
 \* 2582: contig of 2311 bp in length  
 \* 25882: gap of 100 bp  
 \* 28824: contig of 2942 bp in length  
 \* 28924: gap of 100 bp  
 \* 31619: contig of 2695 bp in length  
 \* 31719: gap of 100 bp  
 \* 34679: contig of 2960 bp in length  
 \* 34779: gap of 100 bp  
 \* 37527: contig of 2748 bp in length  
 \* 34780

```

* 37628 37627: gap of 100 bp
* 37628 40146: contig of 2519 bp in length
* 40147 40246: gap of 100 bp
* 40247 43743: contig of 3497 bp in length
* 43744 43843: gap of 100 bp
* 43844 47133: contig of 3290 bp in length
* 47134 47233: gap of 100 bp
* 47234 51023: contig of 3790 bp in length
* 51024 51123: gap of 100 bp
* 51124 54935: contig of 3812 bp in length
* 54936 55035: gap of 100 bp
* 55036 59553: contig of 4518 bp in length
* 59554 6428: gap of 100 bp
* 6429 6428: contig of 4775 bp in length
* 6429 69211: contig of 4683 bp in length
* 69212 69311: gap of 100 bp
* 69312 72901: contig of 3590 bp in length
* 72902 73001: gap of 100 bp
* 73002 76724: contig of 3723 bp in length
* 76725 76824: gap of 100 bp
* 76825 81179: contig of 4355 bp in length
* 81180 81279: gap of 100 bp
* 81280 87009: contig of 5730 bp in length
* 87010 87109: gap of 100 bp
* 87110 90855: contig of 3746 bp in length
* 90856 90955: gap of 100 bp
* 90956 9820: contig of 5565 bp in length
* 9821 9820: gap of 100 bp
* 9821 10321: contig of 5701 bp in length
* 10322 103421: gap of 100 bp
* 10343 108293: contig of 5872 bp in length
* 108294 108393: gap of 100 bp
* 108394 116689: contig of 8296 bp in length
* 116690 116789: gap of 100 bp
* 116790 128264: contig of 8475 bp in length
* 128265 128364: gap of 100 bp
* 128365 136354: contig of 10990 bp in length
* 136355 136454: gap of 100 bp
* 136455 145833: contig of 9379 bp in length.
FEATURES
    Location/Qualifiers
        1..145833
            /organism="Homo sapiens"
            /mol_type="genomic DNA"
            /db_xref="taxon:9606"
            /chromosome="20"
            /map="20"
            /clone="RP11-712N14"
            /clone_lib="RPC1-11 Human Male BAC"
        1..1214
            /note="assembly_fragment"
        1315..2673
            /note="assembly_fragment"
        2774..4520
            /note="assembly_fragment"
        4621..5961
            /note="assembly_fragment"
        6062..7719
            /note="assembly_fragment"
        7820..9799
            /note="assembly_fragment"
        9900..11434
            /note="assembly_fragment"
        11535..14382
            /note="assembly_fragment"
        14483..17148
            /note="assembly_fragment"
        17249..19239
            /note="assembly_fragment"
        19340..21102
            /note="assembly_fragment"
        21203..23371
            /note="assembly_fragment"

```

```

misc_feature 23472..25782
              /note="assembly_fragment"
misc_feature 25883..28824
              /note="assembly_fragment"
misc_feature 28925..31619
              /note="assembly_fragment"
misc_feature 31720..34679
              /note="assembly_fragment"
misc_feature 34780..37527
              /note="assembly_fragment"
misc_feature 37628..40146
              /note="assembly_fragment"
misc_feature 40247..43743
              /note="assembly_fragment"
misc_feature 43844..47133
              /note="assembly_fragment"
misc_feature 47234..51023
              /note="assembly_fragment"
misc_feature 51124..54935
              /note="assembly_fragment"
              clone_end:17
              vector_side:right"
misc_feature 55036..59553

Query Match          100.0%; Score 26; DB 2; Length 145833;
Best Local Similarity 100.0%; Pred.No. 0.21;
Matches 26; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 CTTCTGGAAGTCTGCCAGTGTCTT 26
   |||||
Db 100020 CTTCTGGAAGTCTGCCAGTGTCTT 99995

RESULT 9
AC129114/c
LOCUS          AC129114.2   177004 bp    DNA    linear    HTG 11-OCT-2002
DEFINITION    Rattus norvegicus clone CH230-251P22, WORKING DRAFT SEQUENCE, 4
unordered pieces.
ACCESSION     AC129114
VERSION       AC129114.2   GI:23814684
KEYWORDS      HTG; HTGS_PHASE1; HTGS_DRAFT; HTGS_FULLTOP.
SOURCE        Rattus norvegicus (Norway rat)
ORGANISM      Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
              Rattus.
REFERENCE     1 (bases 1 to 177004)
AUTHORS      Muzny, D., Marie, J., Metzker, M., Lee, S., Abruzzo, S., Adams, C., Alder, J.,
              Allen, C., Allen, H., Altschuld, S., Amin, A., Anguiano, D.,
              Anyalebechi, V., Aoyagi, A., Ayodeji, M., Baca, E., Baden, H.,
              Baldwin, D., Bandaranaike, D., Barber, M., Barnstead, M., Benahmed, F.,
              Biswal, K., Blair, J., Blankenburg, K., Blyth, P., Brown, M.,
              Bryant, N., Buhay, C., Burch, P., Burrell, K., Calderon, E.,
              Cardenas, V., Carter, K., Cavazos, I., Ceasar, H., Center, A.,
              Chacko, J., Chavez, D., Chen, G., Chen, R., Chen, Y., Chen, Z., Chu, J.,
              Cleveland, C., Cockrell, R., Cox, C., Coyle, M., Cree, A., D'Souza, L.,
              Davila, M., Davis, C., Davy-Carroll, L., De Anda, C., Dederich, D.,
              Delgado, O., Denison, S., Deramo, C., Ding, Y., Dinh, H., Divya, K.,
              Draper, H., Dugan-Rocha, S., Dunn, A., Durbin, K., Duval, B., Eaves, K.,
              Egan, A., Escotto, M., Eugene, C., Evans, C., Falls, T., Fan, G.,
              Fernandez, S., Finley, M., Flagg, N., Forbes, L., Foster, M., Foster, P.,
              Fraser, C., Gabisi, A., Ganta, R., Garcia, A., Garner, T., Garza, M.,
              Gebregeorgis, E., Geer, K., Gill, R., Grady, M., Guerra, W., Guevara, W.,
              Gunaratne, P., Haaland, W., Hamill, C., Hamilton, C., Hamilton, K.,
              Harvey, Y., Havlak, P., Hawes, A., Henderson, N., Hernandez, J.,
              Hernandez, R., Hines, S., Hladun, S., Hodgson, A., Hogues, M.,
              Hollins, B., Howells, S., Hulyk, S., Hume, J., Idlebird, D., Jackson, A.,
              Jackson, L., Jacob, L., Jiang, H., Johnson, B., Johnson, R., Jolivet, A.,
              Karpathy, S., Kelly, S., Khan, Z., King, L., Kovar, C., King, L., Liu, J.,
              Kowis, C., Kraft, C., Lebow, H., Levan, J., Lewis, L., Li, Z., Liu, J.,
              Liu, J., Liu, W., Liu, Y., London, P., Longacre, S., Lopez, J.,
              Lorensuwa, L., Louisegh, H., Lozano, R., Lu, X., Ma, J.,
              Maheshwari, M., Mahindartne, M., Mahmoud, M., Malloy, K., Mangum, A.,

```

Mangum, B., Mapua, P., Martin, K., Martin, R., Martinez, E.,  
 Mawhinney, S., McLeod, M.P., McNeill, T.Z., Meenen, E.,  
 Milosavljevic, A., Miner, G., Minja, E., Montemayor, J., Moore, S.,  
 Morgan, M., Morris, K., Morris, S., Munidasa, M., Murphy, M., Naif, L.,  
 Nankervis, C., Neal, D., Newton, N., Nguyen, N., Norris, S.,  
 Nwackeleneh, O., Okwodu, G., Olarnpunsagoon, A., Pal, S., Parks, K.,  
 Pasternak, S., Paul, H., Perez, A., Perez, L., Pfannkuch, C.,  
 Plopper, F., Poindexter, A., Popovic, D., Primus, E., Pu, L.-L.,  
 Puazo, M., Quiroz, J., Rachlin, E., Reeves, K., Regier, M.A., Reigh, R.,  
 Reilly, B., Reilly, M., Ren, Y., Reuter, M., Richards, S., Riggs, F.,  
 Rives, C., Rodkey, T., Rojas, A., Rose, M., Rose, R., Ruiz, S.,  
 Sanders, M., Saverly, G., Scherer, S., Scott, G., Shatsman, S., Shen, H.,  
 Shetty, J., Shvartsbeyn, A., Sisson, I., Sitter, C.D., Smajs, D.,  
 Sneed, J., Sodergren, E., Song, X.-Z., Sorelle, R., Sosa, J.,  
 Steinkle, M., Strong, R., Sutton, A., Svatek, A., Tabor, P., Taylor, C.,  
 Taylor, T., Thomas, N., Thomas, S., Tingey, A., Trejos, Z., Usmani, K.,  
 Valas, R., Vera, V., Villaseana, D., Waldron, L., Walker, B., Wang, J.,  
 Wang, Q., Wang, S., Warren, J., Warren, R., Wei, X., White, F.,  
 Williams, G., Willson, R., Wlezyk, R., Wooden, H., Worley, K.,  
 Wright, D., Wright, R., Wu, J., Yakub, S., Yen, J., Yoon, L., Yoon, V.,  
 Yu, F., Zhang, J., Zhou, J., Zhou, X., Zhao, S., Dunn, D., von  
 Niederhausern, A., Weiss, R., Smith, D.R., Holt, R.A., Smith, H.O.,  
 Weinstein, G. and Gibbs, R.A.  
 Direct Submission  
 Unpublished  
 2 (bases 1 to 177004)  
 Worley, K.C.  
 Direct Submission  
 Submitted (27-JUL-2002) Human Genome Sequencing Center, Department  
 of Molecular and Human Genetics, Baylor College of Medicine, One  
 Baylor Plaza, Houston, TX 77030, USA  
 3 (bases 1 to 177004)  
 Rat Genome Sequencing Consortium.  
 Direct Submission  
 Submitted (11-OCT-2002) Human Genome Sequencing Center, Department  
 of Molecular and Human Genetics, Baylor College of Medicine, One  
 Baylor Plaza, Houston, TX 77030, USA  
 On Oct 11, 2002 this sequence version replaced gi:21998900.  
 The sequence in this assembly is a combination of BAC based reads  
 and whole genome shotgun sequencing reads assembled using Atlas  
 (http://www.hgsc.bcm.tmc.edu/projects/rat/). Each contig described  
 in the feature table below represents a scaffold in the Atlas  
 assembly (a 'contig-scaffold'). Within each contig-scaffold,  
 individual sequence contigs are ordered and oriented, and separated  
 by sized gaps filled with Ns to the estimated size. The sequence  
 may extend beyond the ends of the clone and there may be sequence  
 contigs within a contig-scaffold that consist entirely of whole  
 genome shotgun sequence reads. Both end sequences and whole genome  
 shotgun sequence only contigs will be indicated in the feature  
 table.

\* This record will be updated with the finished sequence  
 \* as soon as it is available and the accession number will  
 \* be preserved.

\* 1 173251: contig of 173251 bp in length.  
 \* 173252 173351: gap of unknown length  
 \* 173352 174391: contig of 1040 bp in length  
 \* 174392 174491: gap of unknown length  
 \* 174492 175689: contig of 1198 bp in length  
 \* 175690 175789: gap of unknown length  
 \* 175790 177004: contig of 1215 bp in length.

#### FEATURES

Location/Qualifiers  
 source  
 1..177004  
 /organism="Rattus norvegicus"  
 /mol\_type="genomic DNA"  
 /db\_xref="taxon:10116"  
 /clone="CH230-251F22"  
 misc\_feature  
 1..1427  
 /note="wgs\_end\_extension  
 clone\_end:T7"  
 6478..7966  
 /note="wgs\_end\_extension  
 clone\_end:T7"  
 misc\_feature  
 9311..10240  
 /note="clone\_boundary  
 clone\_end:T7  
 site:Mbol  
 end\_sequence:RXABT35TJ"  
 54970..165759  
 /note="clone\_boundary  
 clone\_end:Sp8  
 site:Mbol  
 end\_sequence:RXABT35TV"  
 171562..173251  
 /note="wgs\_end\_extension  
 clone\_end:Sp8"  
 misc\_feature  
 54970..165759  
 /note="clone\_boundary  
 clone\_end:Sp8  
 site:Mbol  
 end\_sequence:RXABT35TV"  
 171562..173251  
 /note="wgs\_end\_extension  
 clone\_end:Sp8"  
 ORIGIN

Query Match 81.5%; Score 21.2; DB 2; Length 177004;  
 Best Local Similarity 88.5%; Pred. No. 35;  
 Matches 23; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

Qy 1 CTTTGTGAAGTCTGCCAGTGTCTT 26  
 |||||  
 Db 95163 CTTTCAGGAATTCGCCAGTGTCTT 95138

#### RESULT 10

AC020636/c  
 LOCUS  
 DEFINITION Homo sapiens 3 BAC RP11-251C9 (Roswell Park Cancer Institute Human  
 BAC Library) complete sequence.  
 ACCSSION AC020636  
 VERSION AC020636.14 GI:22003934  
 KEYWORDS HTG.  
 SOURCE Homo sapiens (human)  
 ORGANISM Homo sapiens  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

#### REFERENCE

AUTHORS  
 Muzny, D.M., Adams, C., Adio-Oduola, B., Ali-ouman, F.R., Allen, C.,  
 Alsbrooks, S.L., Amaratunge, H.C., Are, J.R., Ayale, M., Banks, T.,  
 Barbiana, J., Benton, J., Bimaga, K., Blankenburg, K., Bonnin, D.,  
 Bouck, J., Bowie, S., Brieva, M., Brown, E., Brown, M., Bryant, N.P.,  
 Buha, C., Burch, P., Burkett, C., Burrell, K.L., Byrd, N.C.,  
 Carrott, T.F., Carter, M., Cavazos, S.R., Chacko, J., Chavez, D.,  
 Chen, G., Chen, R., Chen, Z., Chowdhry, I., Christopoulos, C.,  
 Cleveland, C.D., Cox, C., Coyle, M.D., Dathorne, S.R., David, R.,  
 Davila, M.L., Davis, C., Davy-Carroll, L., Dederich, D.A.,  
 Delaney, K.R., Delgado, O., Denn, A.L., Ding, Y., Dinh, H.H.,  
 Douthwaite, K.J., Draper, H., Dugan-Rocha, S., Durbin, K.J.,  
 Earnhart, C., Edgar, D., Edwards, C.C., Elhaj, C., Escotto, M.,  
 Falls, T., Ferraguto, D., Flagg, N., Ford, J., Foster, P., Frantz, P.,  
 Gabisi, A., Gao, J., Garcia, A., Garner, T., Garza, N., Gill, R.,

\* NOTE: Estimated insert size may differ from sequence length  
 (see http://www.hgsc.bcm.tmc.edu/docs/genbank\_draft\_data.html).

\* NOTE: This is a 'working draft' sequence. It currently  
 consists of 4 contigs. The true order of the pieces

\* is not known and their order in this sequence record is  
 arbitrary. Gaps between the contigs are represented as

\* runs of N, but the exact sizes of the gaps are unknown.

Assembly program: Phrap; version 0.990329  
 Consensus quality: 156282 bases at least Q40  
 Consensus quality: 158286 bases at least Q30  
 Consensus quality: 159515 bases at least Q20  
 Estimated insert size: 156592; sum-of-contigs estimation  
 Quality coverage: 8x in Q20 bases; sum-of-contigs estimation

Center: Baylor College of Medicine  
 Center code: BCM  
 Web site: http://www.hgsc.bcm.tmc.edu/  
 Contact: hgsc-help@bcm.tmc.edu  
 Project Information  
 Center project name: GMTI  
 Center clone name: CH230-251F22  
 Summary Statistics

Gorrell, J.H., Guevara, W., Gunaratne, P., Hale, S., Hamilton, K., Harris, C., Harris, K., Hart, M., Havlak, P., Hawes, A., He, X., Hernandez, J., Hernandez, O., Hodgson, A., Hoques, M., Holloway, C., Hollins, B., Homs, F., Howard, S., Huber, J., Hulyk, S., Hume, J., Jackson, L.E., Jacobson, B., Jia, Y., Johnson, R., Jolivet, S., Joudah, S., Karlsson, E., Kelly, S., Khan, U., King, L., Korhavi, J., Kovar, C., Kratovic, J., Kureshi, A., Landry, N., Lead, B., Lewis, L.C., Lewis, L., Li, J., Li, Z., Lichtarge, C., Lieu, C., Liu, J., Liu, W., Louis, H., Lozada, R.J., Lu, X., Lucier, A., Lucier, R., Luna, R., Ma, J., Maheshwari, M., Mapua, P., Martin, R., Martindale, A., Martinez, E., Massey, E., Mawhinney, E., McLeod, M.P., Meador, M., Mei, G., Metzker, M., Miner, G., Miner, Z., Morris, S., Moser, M., Neal, D., Nelson, D., Newton, J., Newton, N., Nguyen, A., Nguyen, N., Nguyen, N., Nickerson, E., Nickenko, S., Oguri, M., Okwuonu, G., Orsagunye, N., Oviedo, R., Pace, A., Payton, B., Peery, J., Perez, L., Peters, L., Pickens, R., Primus, E., Pu, L.L., Quiles, M., Ren, Y., Rives, M., Rojas, A., Rojokan, I., Rolfe, M., Ruiz, S., Saverly, G., Scherer, S., Scott, G., Shen, H., Shooshkari, N., Sisson, I., Sodergren, E., Sonaite, T., Sparks, A., Stanley, H., Stone, H., Sutton, A., Svatek, A., Tabor, P., Tamerisa, A., Tamerisa, K., Tang, H., Tansey, J., Taylor, C., Taylor, T., Telford, B., Thomas, N., Thomas, S., Usmani, K., Vasquez, L., Vera, V., Villalón, D., Vinson, R., Wang, Q., Wang, S., Ward-Moore, S., Warren, R., Washington, C., Watlington, S., Williams, G., Williamson, A., Wleczky, R., Woodson, S., Worley, K., Wu, C., Wu, Y., Wu, Y.F., Zhou, J., Zorrilla, S., Naylor, S.L., Weinstein, G. and Gibbs, R.

Direct Submission  
 Unpublished  
 2 (bases 1 to 179497)  
 Worley, K.C.  
 Direct Submission  
 Submitted (07-JAN-2000) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA  
 3 (bases 1 to 179497)  
 Worley, K.C.  
 Direct Submission  
 Submitted (26-JUL-2002) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA  
 4 (bases 1 to 179497)  
 Worley, K.C.  
 Direct Submission  
 Submitted (30-JUL-2002) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA  
 5 (bases 1 to 179497)  
 Worley, K.C.  
 Direct Submission  
 Submitted (24-JAN-2003) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA  
 6 (bases 1 to 179497)  
 Worley, K.C.  
 Direct Submission  
 Submitted (14-MAR-2003) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA  
 7 (bases 1 to 179497)  
 Worley, K.C.  
 Direct Submission  
 Submitted (15-MAR-2003) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA  
 On Jul 30, 2002 this sequence version replaced gi:21954832.  
 INFORMATION: <http://www.hgsc.bcm.tmc.edu/> or email [gc-help@bcm.tmc.edu](mailto:gc-help@bcm.tmc.edu)

CLONE LENGTH: This sequence does not necessarily represent the entire insert of this clone. Overlapping regions of clones are only sequenced and submitted once, so the sequence for the remainder of the insert may be found in the record for the adjacent clones.

Overlapping clones are noted at the beginning and end of the Features listing.

#### ANNOTATION OF FEATURES:

STSS are identified using ePCR (Genome Res. 7:541-550) searches of a local database that includes entries from dbSTS, GDB, and local mapping efforts.

Repeats are identified using RepeatMasker (A. Smit and P. Green, unpublished.) for Human and Mouse sequences.

Genes and Region of sequence similarity are identified by BLAST (Nuc. Acids Res. 25:3389-3402) similarity (expect < 1e-34) to the EST and cDNA sequences. Genes demonstrate at least two exons flanked by consensus splice sites that maintained sequence continuity across the splice junctions. Sequences that are not identical matches are annotated as similar.

SEQUENCING READ COVERAGE: Sequencing is completed to a minimum standard of double strand coverage with a minimum of 2 clones and 2 reads with no ambiguities or 2 chemistries with a minimum of 2 clones and 3 reads with no ambiguities. If the sequence quality for a region does not meet this standard, it will be indicated in the annotation as Low Coverage.

QUALITY OF INDIVIDUAL BASES: This sequence meets stringent quality standards - estimated error rate less than 1 per 10,000 bases. Reports of lowest quality individual bases and measures of base quality are listed below. Description of the metrics can be found at URL:

<http://www.hgsc.bcm.tmc.edu:8088/quality.info/genbank.annotation.ht> ml.

#### FEATURES

Location/Qualifiers	source
1..179497	
/organism="Homo sapiens"	
/mol_type="genomic DNA"	
/db_xref="taxon:9606"	
/chromosome="3"	
/clone="RP11-251C9"	
complement(1..2007)	
/note="overlaps bases 1..2007 of clone AC108674"	
/function="clone overlap"	
complement(1719..2025)	
/rpt_family="AluX"	
complement(2316..2396)	
/rpt_family="MER91A"	
4464..4739	
/standard_name="SHGC-80096"	
6319..6734	
/rpt_family="MLT2FB"	
7371..7400	
/rpt_family="AT-rich"	
7744..7892	
/rpt_family="MIR"	
7988..8274	
/standard_name="SHGC-144786"	
9667..10430	
/rpt_family="MER21B"	
complement(10508..11039)	
/rpt_family="MLT1E1"	
11100..11266	
/rpt_family="MIR"	
complement(11490..11847)	
/rpt_family="MLT1K"	
12962..13005	
/rpt_family="MIR"	
13873..13953	
/rpt_family="MIR"	
13992..14099	
/rpt_family="MER58B"	
14157..14355	
/rpt_family="MER58B"	
14760..15104	
/rpt_family="L1MC5"	
15370..15397	

```

/rpt_family="AT_rich"
complement(15437..15729)
/rpt_family="AluSc"
complement(15863..16265)
/rpt_family="MT1J"
16558..16711
/rpt_family="MIR"
16714..16885
/standard_names="G54351"
complement(17909..17988)
/rpt_family="MADE1"
complement(20197..20270)
/rpt_family="L2"
21547..21594
/rpt_family="CA)n"
21947..22063
/rpt_family="CA)n"
complement(23769..24118)
/rpt_family="MT1J1"
complement(27498..27630)
/rpt_family="FLAM_C"
27771..27798
/rpt_family="AT_rich"
27838..27866
/rpt_family="CA)n"
29110..29356
/rpt_family="MT1J"
29684..29788
/rpt_family="MIR"

Query Match      81.5%; Score 21.2; DB 9; Length 179497;
Best Local Similarity 88.5%; Pred. No. 35;
Matches 23; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 CCTCTGGAAGTCTCCAGTGCCTT 26
Db 29519 CCATCTGGAAGTCTCCAGTGCCTT 29494

RESULT 11
AC105625/c
LOCUS
DEFINITION
Rattus norvegicus clone CH230-117017, WORKING DRAFT SEQUENCE, 2
unordered pieces.
AC105625
AC105625.4 GI:25075238
HTG: HTGS PHASE1: HTGS DRAFT: HTGS_FULLTOP.
Rattus norvegicus (Norway rat)
Rattus norvegicus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
Rattus.
1 (bases 1 to 259967)
Muzny,D,Marie, Metzker,M, Lee, A, Abramzon,S., Adams,C., Alder,J.,
Allen,C., Allen,H., Alsbrooks,S., Amin,A., Anguiano,D.,
Anyalebechi,V., Ayoyagi,A., Ayodeji,M., Baca,E., Baden,H.,
Baldwin,D., Bandaranaike,D., Barber,M., Barnstead,M., Benahmed,F.,
Biswal,K., Blair,J., Blankenburg,K., Blyth,P., Brown,M.,
Bryant,N., Buhay,C., Burch,P., Burrell,K., Calderon,E.,
Cardenas,V., Carter,K., Cavazos,I., Ceasar,H., Center,A.,
Chacko,J., Chavez,D., Chen,G., Chen,R., Chen,Y., Chen,Z., Chu,J.,
Cleveland,C., Cockrell,R., Cox,C., Coyle,M., Cree,A., D'Souza,L.,
Devilla,M., Davis,C., Davy-Carroll,L., De Anda,C., Dederich,D.,
Delgado,O., Denson,S., Deramo,C., Ding,Y., Dinh,H., Divya,K.,
Draper,H., Dugan-Rocha,S., Dunn,A., Durbin,K., Duval,B., Eaves,K.,
Egan,A., Escoto,M., Eugene,C., Evans,C.A., Falls,T., Fan,G.,
Fernandez,S., Finley,M., Flagg,N., Forbes,L., Foster,M., Foster,P.,
Fraser,C.M., Gabisi,A., Ganta,R., Garcia,A., Garner,T., Garza,M.,
Gebregorgis,E., Geer,K., Gill,R., Grady,M., Guerra,W., Guevara,W.,
Gunaratne,P., Haaland,W., Hamil,C., Hamilton,C., Hamilton,K.,
Harvey,Y., Havlak,P., Hawes,A., Henderson,N., Hernandez,J.,
Hernandez,R., Hines,S., Hladun,S.L., Hodgson,A., Hogues,M.,
Hollins,B., Howells,S., Hulyk,S., Hume,J., Idlebird,D., Jackson,A.,

```

```

Jackson,L., Jacob,L., Jiang,H., Johnson,B., Johnson,R., Jolivet,A.,
Karpathy,S., Kelly,S., Khan,Z., King,L., Kovar,C.,
Kowis,C., Kraft,C.L., Lebow,H., Levan,J., Lewis,L., Li,Z., Liu,J.,
Liu,J., Liu,W., Liu,Y., London,P., Longacre,S., Lopez,J.,
Lorensuhewa,L., Loulsegad,H., Lozano,R.J., Lu,X., Ma,J.,
Maheshwari,M., Mahindartine,M., Mahmoud,M., Malloy,K., Mangum,A.,
Mangum,B., Mapua,P., Martin,K., Martin,R., Martinez,E.,
Mawhney,S., McLeod,M.P., McNeill,T.Z., Meenen,E.,
Milosavljevic,A., Miner,G., Minja,G., Montemayor,J., Moore,S.,
Morgan,M., Morris,K., Morris,S., Muidasa,M., Murphy,M., Naif,L.,
Nankervis,C., Neal,D., Newton,N., Nguyen,N., Norris,S.,
Nwackeleme,C., Okwono,G., Olarnpunsagoon,A., Pal,S., Parks,K.,
Pasternak,S., Paul,H., Perez,A., Perez,L., Pfannkuch,C.,
Plopper,F., Poinexter,A., Popovic,D., Primus,E., Pu,L., L.,
Puazo,M., Quiroz,J., Rachlin,E., Reeves,K., Regier,M.A., Reigh,R.,
Reilly,B., Reilly,M., Ren,Y., Reuter,M., Richards,S., Riggs,P.,
Rivers,C., Rodkey,T., Rojase,A., Rose,M., Rose,R., Ruiz,S.,
Sanders,M., Savary,G., Scherer,S., Scott,G., Shatsman,S., Shen,H.,
Shetty,J., Shvartsbeyn,A., Sisson,I., Sitter,C.D., Smajd,D.,
Sneed,A., Sodergren,E., Song,X.-Z., Sorelle,R., Sosa,J.,
Steimle,M., Strong,R., Sutton,A., Svatek,A., Taber,P., Taylor,C.,
Taylor,T., Thomas,N., Thomas,S., Tingey,A., Trejos,Z., Usmani,K.,
Valas,R., Vera,V., Villasana,D., Waldron,L., Walker,B., Wang,J.,
Wang,Q., Wang,S., Warren,J., Warren,R., Wei,X., White,P.,
Williams,G., Willson,R., Wlezyk,R., Wooden,H., Worley,K.,
Wright,D., Wright,R., Wu,J., Yakub,S., Yen,J., Yoon,D., Yoon,V.,
Yu,P., Zhang,J., Zhou,U., Zhou,X., Zhao,S., Dunn,D., von
Niederhausern,A., Weiss,R., Smith,D.R., Holt,R.A., Smith,H.O.,
Weinstock,G. and Gibbs,R.A.
Direct Submission
Unpublished
2 (bases 1 to 259967)
Worley,K.C.
Direct Submission
Submitted (09-JAN-2002) Human Genome Sequencing Center, Department
of Molecular and Human Genetics, Baylor College of Medicine, One
Baylor Plaza, Houston, TX 77030, USA
3 (bases 1 to 259967)
Rat Genome Sequencing Consortium.
Direct Submission
Submitted (19-NOV-2002) Human Genome Sequencing Center, Department
of Molecular and Human Genetics, Baylor College of Medicine, One
Baylor Plaza, Houston, TX 77030, USA
On Nov 19, 2002 this sequence version replaced gi:23101542.
The sequence in this assembly is a combination of BAC based reads
and whole genome shotgun sequencing reads assembled using Atlas
(http://www.hgsc.bcm.tmc.edu/projects/rat/). Each contig described
in the feature table below represents a scaffold in the Atlas
assembly (a 'contig-scaffold'). Within each contig-scaffold,
individual sequence contigs are ordered and oriented, and separated
by sized gaps filled with Ns to the estimated size. The sequence
may extend beyond the ends of the clone and there may be sequence
contigs within a contig-scaffold that consist entirely of whole
genome shotgun sequence reads. Both end sequences and whole genome
shotgun sequence only contigs will be indicated in the feature
table.
----- Genome Center
Center: Baylor College of Medicine
Center code: BCM
Web site: http://www.hgsc.bcm.tmc.edu/
Contact: hgsc-help@bcm.tmc.edu
----- Project Information
Center project name: GNJS
Center clone name: CH230-117017
----- Summary Statistics
Assembly program: Phrap; version 0.990329
Consensus quality: 218447 bases at least Q40
Consensus quality: 221126 bases at least Q30
Consensus quality: 229550 bases at least Q20
Estimated insert size: 234159; sum-of-contigs estimation
Quality coverage: 7x in Q20 bases; sum-of-contigs estimation
-----
* NOTE: Estimated insert size may differ from sequence length

```

(see [http://www.hgsc.bcm.tmc.edu/docs/Genbank\\_draft\\_data.html](http://www.hgsc.bcm.tmc.edu/docs/Genbank_draft_data.html)).

\* NOTE: This is a 'working draft' sequence. It currently  
 \* consists of 2 contigs. The true order of the pieces  
 \* is not known and their order in this sequence record is  
 \* arbitrary. Gaps between the contigs are represented as  
 \* runs of N, but the exact sizes of the gaps are unknown.  
 \* This record will be updated with the finished sequence  
 \* as soon as it is available and the accession number will  
 \* be preserved.

\* 1 14629: contig of 14629 bp in length  
 \* 14630 14729: gap of unknown length  
 \* 14730 25967: contig of 24528 bp in length.

## FEATURES

Location/Qualifiers

1...259967

/organism="Rattus norvegicus"

/mol\_type="genomic DNA"

/db\_xref="taxon:10116"

/clone="CH230-117017"

misc\_feature

1...1404

/note="wgs\_contig"

## ORIGIN

Query Match 81.5%; Score 21.2; DB 2; Length 259967;  
 Best Local Similarity 88.5%; Pred. No. 34;  
 Matches 23; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 CTTCTGGAGTCTCCAGTGTCTT 26

Db 168681 CTTTCAGGAATCTCCAGTGTCTT 168656

## RESULT 12

AC095338/c

LOCUS

AC095338 Rattus norvegicus clone CH230-114A11, linear HTG 10-MAY-2003  
 \*\*\*, unordered pieces. \*\*\* SEQUENCING IN PROGRESS

ACCESSION

AC095338.6 GI:30522611

VERSION HTG; HTGS\_PHASE1; HTGS\_DRAFT; HTGS\_ENRICHED.

KEYWORDS Rattus norvegicus (Norway rat)

SOURCE

ORGANISM Rattus norvegicus

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;

Rattus.

1 (bases 1 to 261492)

Muzny, D. Marie, Metzker, M. Lee, Abramson, S., Adams, C., Alder, J.,

Allen, C., Allen, H., Albrooks, S., Anin, A., Angiano, D.,

Arvallebechi, V., Aovagi, A., Ayodeji, M., Baca, E., Baden, H.,

Baldwin, D., Bandaranaike, D., Barber, M., Barnstead, M., Benahmed, F.,

Biswal, K., Blair, J., Blankenburg, K., Glyth, P., Brown, M.,

Bryant, N., Buhay, C., Burch, P., Burrell, K., Calderon, E.,

Cardenas, V., Carter, K., Cavazos, I., Cesar, H., Center, A.,

Chacko, J., Chavez, D., Chen, G., Chen, R., Chen, Y., Chen, Z., Chu, J.,

Cleveland, C., Cockrell, R., Cox, C., Coyle, M., Cree, A., D'Souza, L.,

Davila, M.L., Davis, C., Davy-Carroll, L., De Anda, C., Dederich, D.,

Delgado, O., Denson, S., Deramo, C., Ding, Y., Dinh, H., Divya, K.,

Drapet, H., Dugan-Rocha, S., Dunn, A., Durbin, K., Duval, B., Eaves, K.,

Egan, A., Escotto, M., Eugene, C., Evans, C.A., Falls, T., Fan, G.,

Fernandez, S., Finley, M., Flagg, N., Forbes, L., Foster, M., Foster, P.,

Fraser, C.M., Gabisi, A., Ganta, R., Garcia, A., Garner, T., Garza, M.,

Georgiadis, E., Geer, K., Gill, R., Grady, M., Guerra, W.,

Gunnarsson, P., Haaland, W., Hamil, C., Hamilton, C., Hamilton, K.,

Harvey, Y., Havlak, P., Haves, A., Henderson, N., Hernandez, J.,

Hernandez, R., Hines, S., Hladun, S.L., Hodgson, A., Hogues, M.,

Hollins, B., Howells, S., Hu, Y., Hume, J., Idlebird, D., Jackson, A.,

Jackson, L., Jacob, L., Jiang, H., Johnson, B., Johnson, R., Jolivet, A.,

Karpatsky, S., Kelly, S., Kelly, S., Khan, Z., King, L., Kovar, C.,

Kowis, C., Kraft, C.L., Lebow, H., Levan, J., Lewis, L., Li, Z., Liu, J.,

Liu, J., Liu, W., Liu, Y., London, P., Longacre, S., Lopez, J.,

Lorenz, H., Louised, H., Lozano, E., Lu, X., Ma, J.,

Mareshwari, M., Mahindratne, M., Mahmood, M., Malloy, K., Mangum, A.,

Mangum, B., Mapua, P., Martin, K., Martin, R., Martinez, E.,

Mawhinney, S., McLeod, M.P., McNeill, T.Z., Meenen, E.,

Milosavljevic, A., Miner, G., Minja, E., Montemayor, J., Moore, S.,  
 Morgan, W., Morris, K., Morris, S., Munitasa, M., Murphy, M., Nait, L.,  
 Nankervis, C., Neal, D., Newton, N., Nguyen, N., Norris, S.,  
 Nwankwelu, O., Okwuonu, G., Olarnunagoon, A., Pal, S., Parks, K.,  
 Pasternak, S., Paul, H., Perez, A., Perez, L., Pfankuch, C.,  
 Plopper, P., Poindexter, A., Popovic, D., Primus, E., Pu, L.,  
 Puazo, M., Quiroz, J., Rachlin, E., Reeves, K., Regier, M.A., Reigh, R.,  
 Reilly, B., Reilly, M., Ren, Y., Reuter, M., Richards, S., Riggs, P.,  
 Rives, C., Rodkey, T., Rojas, A., Rose, M., Rose, R., Ruiz, S.J.,  
 Sanders, W., Savary, G., Scherer, S., Scott, G., Shatsman, S., Shen, H.,  
 Shetty, J., Shvartsbeyn, A., Sisson, I., Sitter, C.D., Smajs, D.,  
 Sneed, A., Sodergren, E., Song, X.-Z., Sorelle, R., Sosa, J.,  
 Steimle, M., Strong, R., Sutton, A., Svatek, A., Tabor, P., Taylor, C.,  
 Taylor, T., Thomas, N., Thomas, S., Tingey, A., Trejos, Z., Umani, K.,  
 Valas, R., Vera, V., Villasana, D., Waldron, L., Walker, B., Wang, J.,  
 Wang, Q., Wang, S., Warren, J., Warren, R., Wei, X., White, F.,  
 Williams, G., Willson, R., Wlezyk, R., Wooden, H., Worley, K.,  
 Wright, D., Wright, R., Wu, J., Yakub, S., Yen, J., Yoon, V.,  
 Yu, F., Zhang, J., Zhou, J., Zhou, X., Zhao, S., Dunn, D., von  
 Niederhausern, A., Weiss, R., Smith, D.R., Holt, R.A., Smith, H.O.,  
 Weinstock, G. and Gibbs, R.A.

Direct Submission

Unpublished

2 (bases 1 to 261492)

Worley, K.C.

Direct Submission

Submitted (16-SEP-2001)

Human Genome Sequencing Center, Department

of Molecular and Human Genetics, Baylor College of Medicine, One

Baylor Plaza, Houston, TX 77030, USA

3 (bases 1 to 261492)

Rat Genome Sequencing Consortium.

Direct Submission

Submitted (10-MAY-2003)

Human Genome Sequencing Center, Department

of Molecular and Human Genetics, Baylor College of Medicine, One

Baylor Plaza, Houston, TX 77030, USA

On May 10, 2003 this sequence version replaced gi:230956505.

The sequence in this assembly is a combination of BAC based reads

and whole genome shotgun sequencing reads assembled using Atlas

(<http://www.hgsc.bcm.tmc.edu/projects/rat/>). Each contig described

in the feature table below represents a scaffold in the Atlas

assembly (a 'contig-scaffold'). Within each contig-scaffold,

individual sequence contigs are ordered and oriented, and separated

by sized gaps filled with Ns to the estimated size. The sequence

may extend beyond the ends of the clone and there may be sequence

contigs within a contig-scaffold that consist entirely of whole

genome shotgun sequence reads. Both end sequences and whole genome

shotgun sequence only contigs will be indicated in the feature

table.

----- Genome Center

Center: Baylor College of Medicine

Center code: BCM

Web site: <http://www.hgsc.bcm.tmc.edu/>

Contact: hgsc-help@bcm.tmc.edu

----- Project Information

Center project name: GGNW

Center clone name: CH230-114A11

----- Summary Statistics

Assembly program: Atlas 3.0;

Consensus quality: 244697 bases at least Q40

Consensus quality: 245889 bases at least Q30

Consensus quality: 251288 bases at least Q20

Estimated insert size: 255194; sum-of-contigs estimation

Quality coverage: 6x in Q20 bases; sum-of-contigs estimation

-----

\* NOTE: Estimated insert size may differ from sequence length

(see [http://www.hgsc.bcm.tmc.edu/docs/genbank\\_draft\\_data.html](http://www.hgsc.bcm.tmc.edu/docs/genbank_draft_data.html))

\* NOTE: This sequence may represent more than one clone.

\* NOTE: This is a 'working draft' sequence. It currently

consists of 7 contigs. The true order of the pieces

is not known and their order in this sequence record is

arbitrary. Gaps between the contigs are represented as

runs of N, but the exact sizes of the gaps are unknown.

\* This record will be updated with the finished sequence

\* as soon as it is available and the accession number will  
\* be preserved.

```

1 249471: contig of 249471 bp in length
2 249571: gap of unknown length
3 250636: contig of 1065 bp in length
4 250737: gap of unknown length
5 250737: contig of 1138 bp in length
6 251875: gap of unknown length
7 251875: contig of 1919 bp in length
8 253894: gap of unknown length
9 253894: contig of 1321 bp in length
10 255315: gap of unknown length
11 255415: contig of 1318 bp in length
12 256733: gap of unknown length
13 256833: contig of 4660 bp in length.

```

## FEATURES

```

source
1..261492
  /organism="Rattus norvegicus"
  /mol_type="genomic DNA"
  /db_xref="taxon:10116"
  /clone="CH230-114All"
1..11458
  /note="wgs_contig"
  /complement(247382..248063)
  /note="clone boundary"
  clone_end:77
  site:EcoRI
  end_sequence:BH288027"

```

## misc\_feature

## misc\_feature

## ORIGIN

```

Query Match      81.5%; Score 21.2; DB 2; Length 261492;
Best Local Similarity 88.5%; Pred. No. 34;
Matches 23; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

```

QY 1 CTTCTGCAAGTCCTCCAGTGCCTT 26

Db 52710 CTTTCAGGAATCTCCAGTGCCTT 52685

## RESULT 13

## AC021312

LOCUS AC021312 76332 bp DNA linear HTG 13-JUL-2000  
Homo sapiens chromosome 4 clone RP11-197L13 map 4, LOW-PASS

## SEQUENCE SAMPLING.

## AC021312.2 GI:9126405

## HTG; HTGS PHASE0.

## Homo sapiens (human)

## Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

1 (bases 1 to 76332)

Barren,B., Linton,L., Nusbaum,C. and Lander,E.

Homo sapiens chromosome 4, clone RP11-197L13

## Unpublished

## 2 (bases 1 to 76332)

Barren,B., Linton,L., Nusbaum,C., Lander,E., Abraham,H., Allen,N.,  
Anderson,S., Baldwin,J., Barna,N., Beckerly,R., Beda,F.,  
Boguslavskiy,L., Boukhgalter,B., Brown,A., Burkett,G., Castle,A.,  
Choepeil,X., Collangelo,M., Collins,S., Collymore,A., Cooke,P.,  
DeArelano,K., Dewar,K., Domino,M., Doyle,M., Fenestor,J.,  
Ferreira,P., FitzHugh,W., Forrest,C., Gage,D., Galagan,J.,  
Gardyna,S., Grant,C., Hagos,B., Headford,A., Horton,L.,  
Howland,J.C., Johnson,R., Jones,C., Kann,L., Karatas,A., Klein,J.,  
Landers,T., Lehoczy,R., Levine,R., Lieu,C., Liu,G., Locke,K.,  
Macdonald,P., Marquis,N., McEwan,P., McGurk,A., McKernan,K.,  
McPheters,R., Meldrum,J., Meneus,L., Morrow,J., Naylor,J.,  
Norman,C.H., O'Connor,T., O'Donnell,P., Olivar,T.M., Peterson,K.,  
Pierre,N., Pisan,C., Pollara,V., Raymond,C., Riley,R., Rothman,D.,  
Roy,A., Santos,R., Severy,P., Spencer,B., Stange-Thomann,N.,  
Stojanovic,N., Subramanian,A., Talamas,J., Testaye,S., Theodore,J.,  
Tirrell,A., Vassiliev,H., Viel,R., Vo,A., Wu,X., Wyman,D., Ye,W.J.,  
Zimmer,A. and Zody,M.

## REFERENCE

## AUTHORS

## TITLE

## JOURNAL

## REFERENCE

## AUTHORS

## Direct Submission

Submitted (16-JAN-2000) Whitehead Institute/MIT Center for Genome  
Research, 320 Charles Street, Cambridge, MA 02141, USA  
On Jul 13, 2000 this sequence version replaced gi:6705794.

All repeats were identified using RepeatMasker:

Smit, A.F.A. & Green, P. (1996-1997)

http://ftp.genome.washington.edu/RM/RepeatMasker.html

----- Genome Center

Center: Whitehead Institute/ MIT Center for Genome Research

Center code: WIBR

Web site: http://www-seq.wi.mit.edu

Contact: sequence.submissions@genome.wi.mit.edu

----- Project Information

Center project name: L4256

Center Clone name: 197\_L13

\* NOTE: This record contains 84 individual  
\* sequencing reads that have not been assembled into  
\* contigs. Runs of N are used to separate the reads  
\* and the order in which they appear is completely  
\* arbitrary. Low-pass sequence sampling is useful for  
\* identifying clones that may be gene-rich and allows  
\* overlap relationships among clones to be deduced.  
\* However, it should not be assumed that this clone  
\* will be sequenced to completion. In the event that  
\* the record is updated, the accession number will  
\* be preserved.

```

1 809: contig of 809 bp in length
2 909: gap of 100 bp
3 910 1713: contig of 804 bp in length
4 1714 1813: gap of 100 bp
5 1814 2627: contig of 814 bp in length
6 2628 2727: gap of 100 bp
7 2728 3523: contig of 796 bp in length
8 3524 3623: gap of 100 bp
9 3624 4432: contig of 809 bp in length
10 4433 5324: gap of 100 bp
11 5325 5424: contig of 792 bp in length
12 5425 6207: contig of 783 bp in length
13 6208 7094: contig of 787 bp in length
14 7095 7194: gap of 100 bp
15 7195 8019: contig of 825 bp in length
16 8020 8119: gap of 100 bp
17 8120 8909: contig of 790 bp in length
18 8910 9009: gap of 100 bp
19 9010 9839: contig of 830 bp in length
20 9840 9939: gap of 100 bp
21 9940 10753: contig of 814 bp in length
22 10754 10853: gap of 100 bp
23 10854 11665: contig of 812 bp in length
24 11666 11765: gap of 100 bp
25 11766 12577: contig of 812 bp in length
26 12578 12677: gap of 100 bp
27 12678 12768: gap of 100 bp
28 12768 13495: contig of 817 bp in length
29 13495 13594: gap of 100 bp
30 13595 14385: contig of 790 bp in length
31 14385 14484: gap of 100 bp
32 14484 15300: contig of 816 bp in length
33 15301 15400: gap of 100 bp
34 15401 16201: contig of 801 bp in length
35 16202 16301: gap of 100 bp
36 16302 17092: contig of 791 bp in length
37 17093 17192: gap of 100 bp
38 17193 17983: contig of 791 bp in length
39 17984 18083: gap of 100 bp
40 18084 18890: contig of 807 bp in length
41 18891 18990: gap of 100 bp
42 18991 19817: contig of 827 bp in length
43 19818 19917: gap of 100 bp
44 19918 20737: contig of 820 bp in length
45 20738 20837: gap of 100 bp

```

```
* * 20839 21660: contig of 823 bp in length
* * 21661 21760: gap of 100 bp
* * 21761 22675: contig of 815 bp in length
* * 22676 22675: gap of 100 bp
* * 22676 23566: contig of 791 bp in length
* * 23567 23566: gap of 100 bp
* * 23567 24371: contig of 805 bp in length
* * 24372 24471: gap of 100 bp
* * 24472 25253: contig of 782 bp in length
* * 25254 25353: gap of 100 bp
* * 25354 26176: contig of 823 bp in length
* * 26177 26276: gap of 100 bp
* * 26277 27096: contig of 820 bp in length
* * 27097 27196: gap of 100 bp
* * 27197 28022: contig of 826 bp in length
* * 28023 28122: gap of 100 bp
* * 28123 28937: contig of 815 bp in length
* * 28938 29037: gap of 100 bp
* * 29038 29856: contig of 819 bp in length
* * 29857 29956: gap of 100 bp
* * 29957 30764: contig of 808 bp in length
* * 30765 30864: gap of 100 bp
* * 30865 31672: contig of 808 bp in length
* * 31673 31772: gap of 100 bp
* * 31773 32594: contig of 822 bp in length
* * 32595 33471: contig of 777 bp in length
* * 33472 33571: gap of 100 bp
* * 33572 34419: contig of 848 bp in length
* * 34420 34519: gap of 100 bp
* * 34520 35349: contig of 830 bp in length
* * 35350 35449: gap of 100 bp
* * 35450 36266: contig of 817 bp in length
* * 36267 36366: gap of 100 bp
* * 36367 37180: contig of 814 bp in length
* * 37181 38094: contig of 814 bp in length
* * 38095 38194: gap of 100 bp
* * 38195 39006: contig of 812 bp in length
* * 39007 39106: gap of 100 bp
* * 39107 39915: contig of 809 bp in length
* * 39916 40015: gap of 100 bp
* * 40016 40825: contig of 810 bp in length
* * 40826 40925: gap of 100 bp
* * 40927 41833: contig of 808 bp in length
* * 41834 42620: contig of 787 bp in length
* * 42621 42720: gap of 100 bp
* * 42721 43510: contig of 790 bp in length
* * 43511 43610: gap of 100 bp
* * 43611 44398: contig of 786 bp in length
* * 44397 44496: gap of 100 bp
* * 44497 45321: contig of 825 bp in length
* * 45322 45421: gap of 100 bp
* * 45422 46237: contig of 816 bp in length
* * 46238 46337: gap of 100 bp
* * 46338 47160: contig of 823 bp in length
* * 47161 47260: gap of 100 bp
* * 47261 48063: contig of 803 bp in length
* * 48064 48163: gap of 100 bp
* * 48164 48971: contig of 808 bp in length
* * 48972 49071: gap of 100 bp
* * 49072 50001: contig of 830 bp in length
* * 50002 50794: contig of 793 bp in length
* * 50795 50894: gap of 100 bp
* * 50895 51725: contig of 831 bp in length
* * 51726 51825: gap of 100 bp
* * 51826 52619: contig of 794 bp in length
* * 52620 52719: gap of 100 bp
* * 52720 53496: contig of 777 bp in length
* * 53497 53595: gap of 100 bp
* * 53596 54412: contig of 816 bp in length
```

```
* * 54413 54512: gap of 100 bp
* * 54513 55337: contig of 825 bp in length
* * 55338 56254: contig of 817 bp in length
* * 56255 57160: contig of 806 bp in length
* * 57161 57260: gap of 100 bp
* * 57261 58081: contig of 821 bp in length
* * 58082 58181: gap of 100 bp
* * 58182 58996: contig of 815 bp in length
* * 58997 59096: gap of 100 bp
* * 59097 59996: contig of 800 bp in length
* * 59997 60802: contig of 806 bp in length
* * 60803 61690: contig of 788 bp in length
* * 61691 61790: gap of 100 bp
* * 61791 62590: contig of 800 bp in length
* * 62591 63480: contig of 790 bp in length
* * 63481 64395: contig of 815 bp in length
* * 64396 64395: contig of 815 bp in length
```

Query Match 80.0%; Score 20.8; DB 2; Length 76332;  
Best Local Similarity 91.7%; Pred. No. 59;  
Matches 22; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 2 CTTCTGGAAGTCTGCCAGTGTCT 25

Db 70153 CTTCTGGAAGTCTGCCAGTGTCT 70176

RESULT 14  
AC011405/c

LOCUS AC011405 184541 bp DNA linear PRI 27-FEB-2002  
DEFINITION Homo sapiens chromosome 5 clone CUB-46B19, complete sequence.  
ACCESSION AC011405

VERSION AC011405.6 GI:18921282

KEYWORDS HTG.

SOURCE Homo sapiens (human)

ORGANISM Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

1 (bases 1 to 184541)

DOE Joint Genome Institute and Stanford Human Genome Center.

Direct Submission

Unpublished

2 (bases 1 to 184541)

DOE Joint Genome Institute.

Direct Submission

Submitted (06-OCT-1999) Production Sequencing Facility, DOE Joint  
Genome Institute, 2800 Mitchell Drive, Walnut Creek, CA 94598, USA

3 (bases 1 to 184541)

DOE Joint Genome Institute and Stanford Human Genome Center.

Direct Submission

Submitted (26-FEB-2002) DOE Joint Genome Institute, 2800 Mitchell  
Drive, Walnut Creek, CA 94598, USA

4 (bases 1 to 184541)

DOE Joint Genome Institute and Stanford Human Genome Center.

Direct Submission

Submitted (27-FEB-2002) DOE Joint Genome Institute, 2800 Mitchell  
Drive, Walnut Creek, CA 94598, USA

On Feb 26, 2002 this sequence version replaced gi:8576061.

Draft Sequence Produced by DOE Joint Genome Institute

www.jgi.doe.gov

Finishing Completed at Stanford Human Genome Center

www.shgc.stanford.edu

Quality: Phrap Quality >=40 100% of Sequence;

Estimated Total Number of Errors is 0.

NOTE: BACTERIAL TRANSPOSON excised at 162460

NOTE: This insert is not the entire sequence of the clone (entire  
sequence is 195.7kb). It is clipped at the overlap with AC034243.

The number of bases overlapped is 32318.



## FEATURES

source  
1. .184541  
/organism="Homo sapiens"  
/mol\_type="genomic DNA"  
/db\_xref="taxon:9606"  
/chromosome="5"  
/clone="CIB-46B19"

misc\_feature  
39283..39323  
/note="Sequence from AC025038 sequenced by GTC."

/chromosome="10"  
/clone="RP11-55C5"  
/clone\_lib="RPC1-11.1"  
39283..39323  
/note="Sequence from AC025038 sequenced by GTC."

## ORIGIN

## ORIGIN

Query Match 80.0%; Score 20.8; DB 9; Length 184541;  
Best Local Similarity 91.7%; Pred. No. 54;  
Matches 22; Conservative 0; Mismatches 2; Indels 0; Gaps 0;  
QY 2 CTCTCGAGAGTCGCCAGTGCTCT 25  
|||||  
Db 108454 CTCTCGAGAGTCGCCAGTGCTT 108431  
|||||

Query Match 78.5%; Score 20.4; DB 9; Length 120194;  
Best Local Similarity 95.5%; Pred. No. 87;  
Matches 21; Conservative 0; Mismatches 1; Indels 0; Gaps 0;  
QY 2 CTCTCGAGAGTCGCCAGTGTC 23  
|||||  
Db 71803 CTCTCGAGAGTCGCCAGTGCC 71782  
|||||

Search completed: February 19, 2004, 23:23:23  
Job time : 119.831 secs

## RESULT 15

## AL355474/c

LOCUS Human DNA sequence from clone Rp11-55C5 on chromosome 10, complete  
DEFINITION sequence.  
ACCESSION AL355474  
VERSION AL355474.15 GI:18121463  
KEYWORDS HTG.  
SOURCE Homo sapiens (human)  
ORGANISM Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
1 (bases 1 to 120194)  
Almeida,J.  
Direct Submission  
Submitted (09-JAN-2002) Wellcome Trust Sanger Institute, Hinxton,  
Cambridgeshire, CB10 1SA, UK. E-mail enquiries:  
humquerry@sanger.ac.uk Clone requests: clonerequest@sanger.ac.uk  
On Jan 10, 2002 this sequence version replaced gi:18072474.  
During sequence assembly data is compared from overlapping clones.  
Where differences are found these are annotated as variations  
together with a note of the overlapping clone name. Note that the  
variation annotation may not be found in the sequence submission  
corresponding to the overlapping clone, as we submit sequences with  
only a small overlap as described above.  
This sequence was finished as follows unless otherwise noted: all  
regions were either double-stranded or sequenced with an alternate  
chemistry or covered by high quality data (i.e. phred quality >=  
30); an attempt was made to resolve all sequencing problems, such  
as compressions and repeats; all regions were covered by at least  
one plasmid subclone or more than one M13 subclone; and the  
assembly was confirmed by restriction digest. The following  
abbreviations are used to associate primary accession numbers given  
in the feature table with their source databases: Em, EMBL; Sw,  
SWISSPROT; Tr, TrEMBL; Wp, WORMPEP; Information on the WORMPEP  
database can be found at  
http://www.sanger.ac.uk/Projects/C\_elegans/wormpep This sequence  
was generated from part of bacterial clone contigs of human  
Chromosome 10, constructed by the Sanger Centre Chromosome 10  
Mapping Group. Further information can be found at  
http://www.sanger.ac.uk/HGP/Chr10  
Rp11-55C5 is from the library RPC1-11.1 constructed by the group of  
Pieter de Jong. For further details see  
http://www.chori.org/bacpac/home.htm  
VECTOR: pBACe3.6

## REFERENCE

## AUTHORS

## TITLE

## JOURNAL

## COMMENT

## FEATURES

## source

1. .120194  
/organism="Homo sapiens"  
/mol\_type="genomic DNA"  
/db\_xref="taxon:9606"

IMPORTANT: This sequence is not the entire insert of clone  
Rp11-55C5 it may be shorter because we sequence overlapping  
sections only once, except for a short overlap.  
The true left end of clone Rp11-55C5 is at 1 in this sequence. The  
true left end of clone Rp11-357A18 is at 118195 in this sequence.

## Location/Qualifiers

1. .120194

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: February 19, 2004, 21:15:48 ; Search time 24.0879 Seconds  
(without alignments)  
4585.415 Million cell updates/sec

Title: US-09-939-853A-141  
Perfect score: 26  
Sequence: 1 cctctggaagtcgcagtgccctt 26

Scoring table: IDENTITY\_NUC  
Gapop 10.0 , Gapext 1.0

Searched: 337863 seqs, 2124099041 residues

Total number of hits satisfying chosen parameters: 6747726

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : N Geneseq 29Jan04: \*  
1: Geneseqn1980s: \*  
2: Geneseqn1980s: \*  
3: Geneseqn2000s: \*  
4: Geneseqn2001as: \*  
5: Geneseqn2001bs: \*  
6: Geneseqn2002as: \*  
7: Geneseqn2003as: \*  
8: Geneseqn2003bs: \*  
9: Geneseqn2003cs: \*  
10: Geneseqn2004s: \*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	26	100.0	26	6	ABK61506 Human NOV
2	26	100.0	444	6	ABQ98669 Human ORF
3	26	100.0	445	5	AAS74747 DNA encod
4	26	100.0	763	6	ABQ98670 Human ORF
5	26	100.0	1183	6	ABK61465 Human CDN
6	26	100.0	2567	6	AAD43980 Human Src
7	18.8	72.3	258	8	ACC58334 Human abr
8	18.8	72.3	258	8	ACC58338 Human abr
9	18.8	72.3	258	8	ACC58337 Mouse abr
10	18.8	72.3	258	8	ACC58336 Human abr
11	18.8	72.3	264	6	AAD27075 Human uro
12	18.8	72.3	288	6	AAD27083 Human uro
13	18.8	72.3	300	1	AAN81299 Human pro
14	18.8	72.3	405	6	AAD27078 Human uro
15	18.8	72.3	429	6	AAD27082 Human uro
16	18.8	72.3	465	7	AAL55230 Tumour an
17	18.8	72.3	465	7	AAL55236 Tumour an
18	18.8	72.3	482	7	ABK62674 Rat seque
19	18.8	72.3	482	9	ADB56463 Toxicity
20	18.8	72.3	482	9	ADB50966 Primary r
21	18.8	72.3	618	9	ADD35169 Mouse mit
22	18.8	72.3	624	2	AAT75153 Metastasi
23	18.8	72.3	645	2	AAT75154 Metastasi

## ALIGNMENTS

RESULT 1									
ABK61506									
ID	ABK61506	standard; DNA; 26 BP.							
XX	XX								
AC	ABK61506;								
DT	18-JUN-2002	(first entry)							
XX	XX								
DE	Human NOV13	RT-PCR probe.							
XX	XX								
KW	Human; ss; NOVX; gene therapy; cardiomyopathy; atherosclerosis;								
KW	cell signal processing disorder; metabolic pathway modulation disorder;								
KW	diabetes; cancer; adenocarcinoma; lymphoma; prostate cancer; probe;								
KW	uterus cancer; immune response; graft-versus-host disease; Exon linking;								
KW	acquired immunodeficiency syndrome; AIDS; asthma; Crohn's disease;								
KW	hypertension; congenital heart defects; multiple sclerosis; inflammation;								
KW	Albright hereditary osteodystrophy; reverse transcriptase PCR.								
XX	XX								
OS	Homo sapiens.								
XX	XX								
PN	WQ200216599-A2.								
XX	XX								
PD	28-FEB-2002.								
XX	XX								
PF	27-AUG-2001; 2001WO-US026510.								
XX	XX								
PR	25-AUG-2000; 2000US-0228191P.								
PR	08-FEB-2001; 2001US-0267300P.								
PR	20-FEB-2001; 2001US-0269961P.								
PR	20-MAR-2001; 2001US-0277337P.								
XX	XX								
PA	(CURA-) CURAGEN CORP.								
PA	(CORT-) COR THERAPEUTICS INC.								
XX	XX								
PI	Burgess CE, Conley PB, Grosse WM, Hart M, Kekuda R, Shimkets RA;								
XX	Spytek KA, Szekeres ES, Tomlinson JE, Topper JN, Yang R;								
XX	WPI; 2002-280937/32.								
DR	DR								
XX	XX								
PT	New polypeptides for treating or preventing a disorder associated with								
PT	them, in humans, e.g. cardiomyopathy, atherosclerosis or cancers.								
XX	XX								
PS	Example 2; Page 234; 263pp; English.								
XX	XX								
CC	The invention relates to an isolated polypeptide (NOVX) a mature form of								
CC	NOVX, a NOVX variant (differing by no more than 15%), the nucleotide								

Aat75155 Metastasi  
Abi99589 Mouse isc  
Aat61672 CDNA enco  
Aat61673 Human pro  
Aaq55771 Pro-uroki  
Aat61674 Human pro  
Aad27080 Human uro  
Aat18237 Pro-uroki  
Aat68797 Coding se  
Aaq06133 Sequence  
Aaq06134 Sequence  
Aaq06135 Sequence  
Aaq04486 Plasmid p  
Aaq04485 Plasmid p  
Aat61671 Human nat  
Aaq10168 Encodes p  
Aaq10169 Encodes p  
Aaq10170 Encodes p  
Aaq23009 Pro-UK. 3  
Aaq41450 Mutant hu  
Aaq48228 PUK Gene.  
Aad27077 Human uro

24 18.8 72.3 666 2 AAT75155  
25 18.8 72.3 689 6 ABI99589  
26 18.8 72.3 1137 2 AAT61672  
27 18.8 72.3 1170 2 AAT61673  
28 18.8 72.3 1188 2 Aaq55771  
29 18.8 72.3 1206 2 Aat61674  
30 18.8 72.3 1212 6 Aad27080  
31 18.8 72.3 1233 2 Aat18237  
32 18.8 72.3 1233 2 Aat68797  
33 18.8 72.3 1236 2 Aaq06133  
34 18.8 72.3 1236 2 Aaq06134  
35 18.8 72.3 1236 2 Aaq06135  
36 18.8 72.3 1236 2 Aaq04486  
37 18.8 72.3 1236 2 Aaq04485  
38 18.8 72.3 1236 2 Aat61671  
39 18.8 72.3 1236 2 Aaq10168  
40 18.8 72.3 1236 2 Aaq10169  
41 18.8 72.3 1236 2 Aaq10170  
42 18.8 72.3 1236 2 Aaq23009  
43 18.8 72.3 1236 2 Aaq41450  
44 18.8 72.3 1236 2 Aaq48228  
45 18.8 72.3 1236 6 Aad27077

CC encoding NOVX (or its complement, fragment or variant). NOVX is NOV1-14,  
 CC 15a, 15b, 16a, and 16b. The NOVX polypeptide, nucleic acid encoding it  
 CC and antibody against it, are useful for treating or preventing (e.g. by  
 CC gene therapy) a NOVX-associated disorder in humans, e.g. cardiomyopathy,  
 CC atherosclerosis, a disorder related to cell signal processing and  
 CC metabolic pathway modulation, diabetes or cancers. The NOVX polypeptide  
 CC and nucleic acids are also useful for determining the presence of  
 CC predisposition to the diseases. The NOVX nucleic acid and polypeptide are  
 CC especially useful in therapeutic or prophylactic applications for  
 CC disorders associated with aberrant NOVX expression or activity, e.g.  
 CC cancers (e.g. adenocarcinoma, lymphoma, prostate cancer or uterus  
 CC cancer), immune response, graft-versus-host disease, acquired  
 CC immunodeficiency syndrome (AIDS), asthma, Crohn's disease, hypertension,  
 CC congenital heart defects, multiple sclerosis, inflammation or Albright  
 CC hereditary osteodysplasia and many other diseases listed in the  
 CC specification. The DNA encoding the protein is useful in gene therapy for  
 CC treating the conditions. This is also useful in detection assays,  
 CC chromosome mapping, tissue typing, diagnostic or prognostic assays, or  
 CC for developing a powerful assay system for functional analysis of various  
 CC human disorders, as well as in diagnostic applications. The present  
 CC sequence is a reverse transcriptase (RT)-PCR probe used to measure tissue  
 CC specific expression of mRNA encoding a NOVX protein  
 XX  
 SQ Sequence 26 BP; 3 A; 8 C; 6 G; 9 T; 0 U; 0 Other;

Query Match 100.0%; Score 26; DB 6; Length 26;  
 Best Local Similarity 100.0%; Pred. No. 0.064;  
 Matches 26; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 CCTTCTGGAAGTCTGCCAGTGTCTT 26  
 Db 1 CCTTCTGGAAGTCTGCCAGTGTCTT 26

## RESULT 2

ABQ98669/c  
 ID ABQ98669 standard; DNA; 444 BP.

AC ABQ98669;

DT 04-NOV-2002 (first entry)

DE Human ORF476 coding sequence.

XX Cytostatic; Cardiant; Anti-allergic; Immunosuppressive; Vulnary;  
 XX Antiinflammatory; gene therapy; human; ORFX; atherogenic; platelet;  
 XX human umbilical vein endothelial cell; HUVEC; atherosclerotic plaque;  
 XX cancer; cardiovascular disease; allergy; autoimmune disease;  
 XX wound healing; blood coagulation disorder; inflammatory disorder; ds.

OS Homo sapiens.

XX US2002082206-A1.

DN 27-JUN-2002.

XX 30-MAY-2001; 2001US-00867550.

XX 30-MAY-2000; 2000US-0208427P.

XX (LEAC/) LEACH M D.

PA (MEHR/) MEHRABAN F.

PA (CONL/) CONLEY P B.

PA (TOPP/) TOPPER J N.

PA (LAWD/) LAW D.

PI Leach MD, Mehraban F, Conley PB, Topper JN, Law D;

XX WPI; 2002-626554/67.

DR P-PSDB; ABP64106.

XX New polypeptide designated ORFX are present in human atherogenic cells  
 PT and are useful to prevent and treat ORFX-associated disorders including

PT cancer, allergy, wound healing or autoimmune, cardiovascular or  
 PT inflammatory disease.

XX Claim 2; SEQ ID NO 951; 78pp; English.

XX The present invention relates to novel human ORFX polypeptides and their  
 CC coding sequences (ABP63631-ABP64681 and ABQ98194-ABQ99267). The sequences  
 CC were discovered in human atherogenic cells, in particular in platelets  
 CC and human umbilical vein endothelial cells (HUVEC) and are expressed in  
 CC many other tissues as well. Atherogenic cells are cells which have the  
 CC potential to develop atherosclerotic plaques. The ORFX polypeptides and  
 CC nucleic acids are useful for treating or preventing a pathological  
 CC condition associated with an ORFX-associated disorder, e.g. cancer,  
 CC cardiovascular disease, allergy, autoimmune disease, wound healing, blood  
 CC coagulation disorders or inflammatory disorders. Note: The sequence data  
 CC for this patent did not form part of the printed specification, but was  
 CC obtained in electronic format directly from the USPTO web site at  
 CC seqdata.uspto.gov/sequence.html?docID=20020082206

XX Sequence 444 BP; 103 A; 128 C; 132 G; 81 T; 0 U; 0 Other;

Query Match 100.0%; Score 26; DB 6; Length 444;

Best Local Similarity 100.0%; Pred. No. 0.09;

Matches 26; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 CCTTCTGGAAGTCTGCCAGTGTCTT 26

Db 35 CCTTCTGGAAGTCTGCCAGTGTCTT 10

## RESULT 3

AAS74747

ID AAS74747 standard; cDNA; 445 BP.

AC AAS74747;

DT 13-FEB-2002 (first entry)

XX DNA encoding novel human diagnostic protein #10551.

XX Human; chromosome mapping; gene mapping; gene therapy; forensic;  
 KW food supplement; medical imaging; diagnostic; genetic disorder; ss.

XX Homo sapiens.

XX WO200175067-A2.

XX 11-OCT-2001.

XX 30-MAR-2001; 2001WO-US008631.

XX 31-MAR-2000; 2000US-00540217.

XX 23-AUG-2000; 2000US-00649167.

XX (HYSE-) HYSEQ INC.

XX Drmanac RT, Liu C, Tang YT;

XX WPI; 2001-639362/73.

XX P-PSDB; ABG10560.

XX New isolated polynucleotide and encoded polypeptides, useful in  
 PT diagnostics, forensics, gene mapping, identification of mutations  
 PT responsible for genetic disorders or other traits and to assess  
 PT biodiversity.

XX Claim 1; SEQ ID NO 10551; 103pp; English.

XX The invention relates to isolated polynucleotide (I) and polypeptide (II)  
 CC sequences. (I) is useful as hybridisation probes, polymerase chain  
 CC reaction (PCR) primers, oligomers, and for chromosome and gene mapping,  
 CC and in recombinant production of (II). The polynucleotides are also used  
 CC in diagnostics as expressed sequence tags for identifying expressed

CC genes. (I) is useful in gene therapy techniques to restore normal  
 CC activity of (II) or to treat disease states involving (II). (II) is  
 CC useful for generating antibodies against it, detecting or quantitating a  
 CC polypeptide in tissue, as molecular weight markers and as a food  
 CC supplement. (II) and its binding partners are useful in medical imaging  
 CC of sites expressing (II). (I) and (II) are useful for treating disorders  
 CC involving aberrant protein expression or biological activity. The  
 CC polypeptide and polynucleotide sequences have applications in  
 CC diagnostics, forensics, gene mapping, identification of mutations  
 CC responsible for genetic disorders or other traits to assess biodiversity  
 CC and to produce other types of data and products dependent on DNA and  
 CC amino acid sequences. AAS64197-AAS94564 represent novel human diagnostic  
 CC coding sequences of the invention. Note: The sequence data for this  
 CC patent did not appear in the printed specification, but was obtained in  
 CC electronic format directly from WIPO at  
 CC ftp.wipo.int/pub/published\_pat\_sequences

SQ Sequence 445 BP; 89 A; 112 C; 143 G; 101 T; 0 U; 0 Other;  
 Query Match 100.0%; Score 26; DB 5; Length 445;  
 Best Local Similarity 100.0%; Pred. No. 0.09;  
 Matches 26; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 CCTTCTGGAAGTCTGCCAGTGTCTT 26  
 Db 193 CCTTCTGGAAGTCTGCCAGTGTCTT 218

RESULT 4  
 ABQ98670/c  
 ID ABQ98670 standard; DNA; 763 BP.

AC ABQ98670;

XX 04-NOV-2002 (first entry)

DE Human ORF477 coding sequence.

XX Cytostatic; Cardiant; Anti-allergic; Immunosuppressive; Vulnervary;  
 KW Antinflammatory; gene therapy; human; ORFX; atherogenic; platelet;  
 KW human umbilical vein endothelial cell; HUVEC; atherosclerotic plaque;  
 KW cancer; cardiovascular disease; allergy; autoimmune disease;  
 KW wound healing; blood coagulation disorder; inflammatory disorder; ds.

OS Homo sapiens.

XX US2002082206-A1.

XX 27-JUN-2002.

XX 30-MAY-2001; 2001US-00867550.

XX 30-MAY-2000; 2000US-0208427P.

XX (LEAC/) LEACH M D.

XX (MEHR/) MEHRABAN F.

XX (CONL/) CONLEY P B.

XX (TOPP/) TOPPER J N.

XX (LAWD/) LAW D.

XX Leach MD, Mehraban F, Conley PB, Topper JN, Law D;

XX WPI; 2002-626554/67.

XX P-PSDB; ABP64107.

XX New polypeptide designated ORFX are present in human atherogenic cells  
 PT and are useful to prevent and treat ORFX-associated disorders including  
 PT cancer, allergy, wound healing or autoimmune, cardiovascular or  
 PT inflammatory disease.

XX Claim 2; SEQ ID NO 953; 78pp; English.

XX The present invention relates to novel human ORFX polypeptides and their

CC coding sequences (ABP63631-ABP64681 and ABQ98194-ABQ99267). The sequences  
 CC were discovered in human atherogenic cells, in particular in platelets  
 CC and human umbilical vein endothelial cells (HUVEC) and are expressed in  
 CC many other tissues as well. Atherogenic cells are cells which have the  
 CC potential to develop atherosclerotic plaques. The ORFX polypeptides and  
 CC nucleic acids are useful for treating or preventing a pathological  
 CC condition associated with an ORFX-associated disorder, e.g. cancer,  
 CC cardiovascular disease, allergy, autoimmune disease, wound healing, blood  
 CC coagulation disorders or inflammatory disorders. Note: The sequence data  
 CC for this patent did not form part of the printed specification, but was  
 CC obtained in electronic format directly from the USPTO web site at  
 CC seqdata.uspto.gov/sequence.html?docID=20020082206

SQ Sequence 763 BP; 176 A; 222 C; 218 G; 147 T; 0 U; 0 Other;

Query Match 100.0%; Score 26; DB 6; Length 763;

Best Local Similarity 100.0%; Pred. No. 0.096;

Matches 26; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 CCTTCTGGAAGTCTGCCAGTGTCTT 26

Db 164 CCTTCTGGAAGTCTGCCAGTGTCTT 139

RESULT 5

ID ABK61465/c

XX ABK61465 standard; cDNA; 1183 BP.

AC ABK61465;

XX 18-JUN-2002 (first entry)

DE Human cDNA encoding protein NOV13.

XX Human; Gene; ss; NOVX; gene therapy; cardiomyopathy; atherosclerosis;  
 KW cell signal processing disorder; metabolic pathway modulation disorder;  
 KW diabetes; cancer; adenocarcinoma; lymphoma; prostate cancer;  
 KW uterus cancer; immune response; graft-versus-host disease;  
 KW acquired immunodeficiency syndrome; AIDS; asthma; Crohn's disease;  
 KW hypertension; congenital heart defects; multiple sclerosis; inflammation;  
 KW Albright hereditary osteodystrophy.

OS Homo sapiens.

XX WO200216599-A2.

XX 28-FEB-2002.

XX 27-AUG-2001; 2001WO-US026510.

XX 25-AUG-2000; 2000US-0228191P.

XX 08-FEB-2001; 2001US-0267300P.

XX 20-FEB-2001; 2001US-0269961P.

XX 20-MAR-2001; 2001US-0277337P.

XX (CURA-) CURAGEN CORP.

XX (CORT-) COR THERAPEUTICS INC.

XX Burgess CE, Conley PB, Grosse WM, Hart M, Kekuda R, Shimkets RA;

XX Spytek KA, Szekeres ES, Tomlinson JE, Topper JN, Yang R;

XX WPI; 2002-280937/32.

XX P-PSDB; AAU91308.

XX New polypeptides for treating or preventing a disorder associated with  
 PT them, in humans, e.g. cardiomyopathy, atherosclerosis or cancers.

XX Claim 1; Page 98; 263pp; English.

XX The invention relates to an isolated polypeptide (NOVX) a mature form of  
 CC NOVX, a NOVX variant (differing by no more than 15%), the nucleotide  
 CC encoding NOVX (or its complement, fragment or variant). NOVX is NOV1-14,  
 CC 15a, 15b, 16a, and 16b. The NOVX polypeptide, nucleic acid encoding it

CC and antibody against it, are useful for treating or preventing (e.g. by  
 CC gene therapy) a NOVX-associated disorder in humans, e.g. cardiomyopathy,  
 CC atherosclerosis, a disorder related to cell signal processing and  
 CC metabolic pathway modulation, diabetes or cancers. The NOVX polypeptide  
 CC and nucleic acids are also useful for determining the presence of  
 CC predisposition to the diseases. The NOVX nucleic acid and polypeptide are  
 CC especially useful in therapeutic or prophylactic applications for  
 CC disorders associated with aberrant NOVX expression or activity, e.g.  
 CC cancers (e.g. adenocarcinoma, lymphoma, prostate cancer or uterus  
 CC cancer), immune response, graft-versus-host disease, acquired  
 CC immunodeficiency syndrome (AIDS), asthma, Crohn's disease, hypertension,  
 CC congenital heart defects, multiple sclerosis, inflammation or Albrit  
 CC hereditary osteodysplasia and many other diseases listed in the  
 CC specification. The DNA encoding the protein is useful in gene therapy for  
 CC treating the conditions. This is also useful in detection assays,  
 CC chromosome mapping, tissue typing, diagnostic or prognostic assays, or  
 CC for developing a powerful assay system for functional analysis of various  
 CC human disorders, as well as in diagnostic applications. The present  
 CC sequence encodes a NOVX protein

SQ Sequence 1183 BP; 251 A; 359 C; 333 G; 240 T; 0 U; 0 Other;

Query Match 100.0%; Score 26; DB 6; Length 1183;  
 Best Local Similarity 100.0%; Pred. No. 0.1;  
 Matches 26; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 CCTCTGGAAGTCGCCAGTGTCTT 26

Db 276 CCTCTGGAAGTCGCCAGTGTCTT 251

# RESULT 6

AA43980/c  
 ID AAD43980 standard; cDNA; 2567 BP.

AC AAD43980;

DT 13-DEC-2002 (first entry)

DE Human Src-Like Adapter Protein-2 (hSLAP-2) cDNA.

XX Human; SH2/SH3-domain-containing adapter; Src-Like Adapter Protein-2;  
 KW SLAP-2; immune disorder; signal transduction; autoimmune disease; cancer;  
 KW neoplasm; inflammation disorder; rheumatoid arthritis; osteoarthritis;  
 KW psoriasis; rhinitis; allergy; inflammatory bowel disease; gene therapy;  
 KW Crohn's disease; systemic lupus erythematosus; tissue/organ rejection;  
 KW multiple sclerosis; asthma; acute respiratory distress syndrome;  
 KW pulmonary disorder; dermatological; neuroprotective; gene; ss.

OS Homo sapiens.

XX Key Location/Qualifiers  
 FT CDS 415..1200  
 FT FT /\*tag= a  
 FT FT /product= "Human SLAP-2"

XX WO200242457-A1.

XX 30-MAY-2002.

XX 20-NOV-2001; 2001WO-US043367.

XX 22-NOV-2000; 2000US-0252545P.

XX (BRIM ) BRISTOL-MYERS SQUIBB CO.

XX Chang H, Yang W, Wu Y, Whitney GS, Perez-Villar JJ, Kanner SB;

XX WPI; 2002-463632/49.

XX P-PSDB; AAE26357.

XX Novel substantially purified human SH2/SH3-domain-containing adapter  
 PT polypeptide, termed Src-Like Adapter Protein-2, useful for therapeutic

PT intervention in immunological and inflammatory disorders and cancer.

XX Claim 2; Fig 1; 85pp; English.

CC The invention relates to a substantially purified human SH2/SH3-domain-  
 CC containing adapter polypeptide, termed Src-Like Adapter Protein-2 (SLAP-  
 CC 2). The invention is useful for treating an immune disorder involving  
 CC hyperactivity of B- or T- lymphocytes in a mammal. The invention is  
 CC useful for screening for antagonists or inhibitors of the interaction of  
 CC hSLAP-2 with cellular signalling compounds, for diagnosing, treating or  
 CC preventing diseases or disorders associated with aberrant or uncontrolled  
 CC cellular signal transduction, for determining those cellular signalling  
 CC molecules which associate with hSLAP-2 and which provide critical signals  
 CC for cell activation, and as effectors in methods to affect T-cell  
 CC activation. The invention is useful in screening assays to identify and  
 CC detect candidate bioactive agents that modulate hSLAP-2 bioactivity, for  
 CC potential use to treat autoimmune diseases which may be caused by  
 CC hyperactivated B cells, as well as to treat diseases which may be caused  
 CC by hyperactivated T cells, in addition to other immune system related  
 CC conditions, diseases, or disorders, T-cell and B-cell neoplasms,  
 CC inflammation disorders, diseases and conditions, rheumatoid arthritis,  
 CC osteoarthritis, psoriasis, rhinitis, inflammatory bowel disease (Crohn's  
 CC and ulcerative colitis), allergies, particularly those involving  
 CC hyperactivity of B-cells and T-cells, or other immune cells, such as  
 CC mast cells or eosinophils, autoimmune diseases such as systemic lupus  
 CC erythematosus and multiple sclerosis, pulmonary diseases including  
 CC asthma, acute respiratory distress syndrome, and chronic obstructive  
 CC pulmonary disorder, tissue/ organ rejection and cancer. The invention is  
 CC useful in gene therapy. The present sequence is human SLAP-2 cDNA

SQ Sequence 2567 BP; 611 A; 741 C; 566 G; 549 T; 0 U; 0 Other;

Query Match 100.0%; Score 26; DB 6; Length 2567;

Best Local Similarity 100.0%; Pred. No. 0.11;

Matches 26; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 CCTCTGGAAGTCGCCAGTGTCTT 26

Db 293 CCTCTGGAAGTCGCCAGTGTCTT 268

# RESULT 7

ACC58334

ID ACC58334 standard; cDNA; 258 BP.

AC ACC58334;

XX 26-AUG-2003 (first entry)

DE Human abrogen (hATF-kringle) coding sequence.

XX Human; abrogen; kringle; hATF-kringle; angiogenesis; inhibitor;  
 KW urokinase; plasminogen activator; tumour; metastasis; cytostatic;  
 KW gene therapy; gene; ss.

OS Homo sapiens.

XX Key Location/Qualifiers  
 FT CDS 1..258  
 FT FT /\*tag= a  
 FT FT /partial  
 FT FT /product= "Human abrogen"  
 FT FT /note= "no start or stop codon"

XX WO2003042354-A2.

XX 22-MAY-2003.

XX 04-SEP-2002; 2002WO-US027885.

XX 04-SEP-2001; 2001US-0316300P.

XX (AVET ) AVENTIS PHARM INC.

XX PI Nesbit M, Fong TC, Brockstedt D;  
 XX DR WPI; 2003-449566/42.  
 XX DR P-PSDB; ABR42599.  
 XX PT New abrogen polypeptide, useful for treating an angiogenesis related  
 XX PT diseases e.g. tumor metastasis.  
 XX PS Claim 32; Page 24-25; 95pp; English.  
 XX CC The present sequence is a coding sequence for a novel human abrogen,  
 CC designated hATF-kringle, comprising the human urokinase plasminogen  
 CC activator kringle domain. Abrogens such as hATF-kringle are potent  
 CC inhibitors of endothelial proliferation and angiogenesis. Abrogen  
 CC polypeptides are capable of inhibiting or reducing cell proliferation  
 CC induced by both basic fibroblast growth factor (bFGF) and vascular  
 CC endothelial growth factor in a specific endothelial cell proliferation  
 CC assay; angiostatin only inhibits bFGF induced proliferation in this  
 CC assay. Vectors that expressed abrogen polypeptides in vivo were shown to  
 CC reduce tumour metastasis in 2 lung cancer models. The invention provides  
 CC abrogen polypeptides and polynucleotides, and methods of using these to  
 CC treat an angiogenesis-related disease or disorder, e.g. tumour metastasis  
 CC (claimed)  
 XX SQ Sequence 258 BP; 63 A; 74 C; 71 G; 50 T; 0 U; 0 Other;  
 Query Match 72.3%; Score 18.8; DB 8; Length 258;  
 Best Local Similarity 90.9%; Pred. No. 1.3e+02;  
 Matches 20; Conservative 0; Mismatches 2; Indels 0; Gaps 0;  
 QY 5 CTGGAAGTCTGCCAGTGTCTT 26  
 Db 78 CTGGAAGTCTGCCAGTGTCTT 99  
 RESULT 8  
 ACC58338  
 ID ACC58338 standard; cDNA; 258 BP.  
 AC ACC58338;  
 XX 26-AUG-2003 (first entry)  
 DT Human abrogen (hATF-kringle) coding sequence.  
 DE Human; abrogen; kringle; hATF-kringle; angiogenesis; inhibitor;  
 KW urokinase plasminogen activator; tumour; metastasis; cytostatic;  
 KW gene therapy; gene; ss.  
 XX Homo sapiens.  
 XX Key Location/Qualifiers  
 XX CDS 1..258  
 XX /\*tag= a  
 XX /partial  
 XX /product= "Human abrogen"  
 XX /note= "no start or stop codon"  
 XX WO2003042354-A2.  
 XX 22-MAY-2003.  
 XX 04-SEP-2002; 2002WO-US027885.  
 XX 04-SEP-2001; 2001US-0316300P.  
 XX (AVET ) AVENTIS PHARM INC.  
 XX Nesbit M, Fong TC, Brockstedt D;  
 XX WPI; 2003-449566/42.  
 XX P-PSDB; ABR42617.

XX PT New abrogen polypeptide, useful for treating an angiogenesis related  
 XX PT diseases e.g. tumor metastasis.  
 XX PS Disclosure; Page 95; 95pp; English.  
 XX CC The present sequence is a coding sequence for a novel human abrogen,  
 CC designated hATF-kringle, comprising the human urokinase plasminogen  
 CC activator kringle domain. Abrogens such as hATF-kringle are potent  
 CC inhibitors of endothelial proliferation and angiogenesis. Abrogen  
 CC polypeptides are capable of inhibiting or reducing cell proliferation  
 CC induced by both basic fibroblast growth factor (bFGF) and vascular  
 CC endothelial growth factor in a specific endothelial cell proliferation  
 CC assay; angiostatin only inhibits bFGF induced proliferation in this  
 CC assay. Vectors that expressed abrogen polypeptides in vivo were shown to  
 CC reduce tumour metastasis in 2 lung cancer models. The invention provides  
 CC abrogen polypeptides and polynucleotides, and methods of using these to  
 CC treat an angiogenesis-related disease or disorder, e.g. tumour metastasis  
 CC (claimed)  
 XX SQ Sequence 258 BP; 61 A; 73 C; 72 G; 50 T; 0 U; 2 Other;  
 Query Match 72.3%; Score 18.8; DB 8; Length 258;  
 Best Local Similarity 90.9%; Pred. No. 1.3e+02;  
 Matches 20; Conservative 0; Mismatches 2; Indels 0; Gaps 0;  
 QY 5 CTGGAAGTCTGCCAGTGTCTT 26  
 Db 78 CTGGAAGTCTGCCAGTGTCTT 99  
 RESULT 9  
 ACC58337  
 ID ACC58337 standard; cDNA; 258 BP.  
 XX ACC58337;  
 XX 26-AUG-2003 (first entry)  
 DT Mouse abrogen (MATF-kringle) coding sequence.  
 DE Mouse; abrogen; kringle; MATF-kringle; angiogenesis; inhibitor;  
 KW urokinase plasminogen activator; tumour; metastasis; cytostatic;  
 KW gene therapy; gene; ss.  
 XX Mus sp.  
 XX Key Location/Qualifiers  
 XX CDS 1..258  
 XX /\*tag= a  
 XX /partial  
 XX /product= "Mouse abrogen"  
 XX /note= "no start or stop codon"  
 XX WO2003042354-A2.  
 XX 22-MAY-2003.  
 XX 04-SEP-2002; 2002WO-US027885.  
 XX 04-SEP-2001; 2001US-0316300P.  
 XX (AVET ) AVENTIS PHARM INC.  
 XX Nesbit M, Fong TC, Brockstedt D;  
 XX WPI; 2003-449566/42.  
 XX P-PSDB; ABR42602.  
 XX New abrogen polypeptide, useful for treating an angiogenesis related  
 XX PT diseases e.g. tumor metastasis.  
 XX PS Claim 32; Page 25-26; 95pp; English.

XX The present sequence is a coding sequence for a novel murine abrogen,  
 CC designated hATF-kringle, comprising the mouse urokinase plasminogen  
 CC activator kringle domain. Abrogens such as hATF-kringle are potent  
 CC inhibitors of endothelial proliferation and angiogenesis. Abrogen  
 CC polypeptides are capable of inhibiting or reducing cell proliferation  
 CC induced by both basic fibroblast growth factor (bFGF) and vascular  
 CC endothelial growth factor in a specific endothelial cell proliferation  
 CC assay; angiotensin only inhibits bFGF induced proliferation in this  
 CC assay. Vectors that expressed abrogen polypeptides in vivo were shown to  
 CC reduce tumour metastasis in 2 lung cancer models. The invention provides  
 CC abrogen polypeptides and polynucleotides, and methods of using these to  
 CC treat an angiogenesis-related disease or disorder, e.g. tumour metastasis  
 CC (claimed)

XX SQ Sequence 258 BP; 62 A; 73 C; 72 G; 51 T; 0 U; 0 Other;

Query Match 72.3%; Score 18.8; DB 8; Length 258;  
 Best Local Similarity 90.9%; Pred. No. 1.3e+02;  
 Matches 20; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 5 CTGGAAGTCTGCCAGTGTCTT 26  
 |||||  
 Db 78 CTGGAAGTCTGCCAGTGTCTT 99

RESULT 10  
 ACC58336  
 ID ACC58336 standard; cDNA; 258 BP.  
 XX  
 AC ACC58336;  
 XX  
 DT 26-AUG-2003 (first entry)  
 XX  
 DE Human abrogen (hATF-kringle) coding sequence.  
 XX  
 KW Human; abrogen; kringle; hATF-kringle; angiogenesis; inhibitor;  
 KW urokinase plasminogen activator; tumour; metastasis; cytostatic;  
 KW Gene therapy; gene; ss.  
 XX  
 OS Homo sapiens.

XX Key Location/Qualifiers  
 FH CDS 1..258  
 FT /\*tag= a  
 FT /partial  
 FT /product= "Human abrogen"  
 FT /note= "no start or stop codon"

XX WO2003042354-A2.  
 XX  
 XX 22-MAY-2003.  
 XX  
 XX 04-SEP-2002; 2002WO-US027885.  
 XX  
 XX 04-SEP-2001; 2001US-0316300P.  
 XX  
 XX (AVET ) AVENTIS PHARM INC.  
 XX  
 XX Nesbit M, Fong TC, Brockstedt D;  
 XX  
 XX WPI; 2003-449566/42.  
 XX P-PSDB; ABR42601.  
 XX  
 XX New abrogen polypeptide, useful for treating an angiogenesis related  
 XX diseases e.g. tumor metastasis.  
 XX  
 XX Claim 32; Page 25; 95pp; English.

XX The present sequence is a coding sequence for a novel human abrogen,  
 CC designated hATF-kringle, comprising the human urokinase plasminogen  
 CC activator kringle domain. Abrogens such as hATF-kringle are potent  
 CC inhibitors of endothelial proliferation and angiogenesis. Abrogen

CC polypeptides are capable of inhibiting or reducing cell proliferation  
 CC induced by both basic fibroblast growth factor (bFGF) and vascular  
 CC endothelial growth factor in a specific endothelial cell proliferation  
 CC assay; angiotensin only inhibits bFGF induced proliferation in this  
 CC assay. Vectors that expressed abrogen polypeptides in vivo were shown to  
 CC reduce tumour metastasis in 2 lung cancer models. The invention provides  
 CC abrogen polypeptides and polynucleotides, and methods of using these to  
 CC treat an angiogenesis-related disease or disorder, e.g. tumour metastasis  
 CC (claimed)

XX SQ Sequence 258 BP; 62 A; 74 C; 72 G; 50 T; 0 U; 0 Other;

Query Match 72.3%; Score 18.8; DB 8; Length 258;  
 Best Local Similarity 90.9%; Pred. No. 1.3e+02;  
 Matches 20; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

QY 5 CTGGAAGTCTGCCAGTGTCTT 26  
 |||||  
 Db 78 CTGGAAGTCTGCCAGTGTCTT 99

RESULT 11  
 AAD27075  
 ID AAD27075 standard; DNA; 264 BP.  
 XX  
 AC AAD27075;  
 XX  
 DT 09-APR-2002 (first entry)  
 XX  
 DE Human urokinase-type plasminogen activator (uPA) kringle DNA.

XX Human; urokinase-type plasminogen activator; uPA; therapy; hypertension;  
 KW stroke; hypertension; atherosclerosis; heart attack; thrombotic disorder;  
 KW microvascular occlusion; angiogenic disorder; pulmonary fibrosis; asthma;  
 KW tumour cell metastasis; glaucoma; diabetic retinopathy; wound healing;  
 KW clotting disorder; uterine contraction disorder; respiratory disease;  
 KW male impotence; adult respiratory distress syndrome; ds.  
 XX  
 OS Homo sapiens.

XX Key Location/Qualifiers  
 FH CDS 1..264  
 FT /\*tag= a  
 FT /product= "Human uPA kringle"  
 FT /note= "CDS does not include start and stop codon"  
 FT /partial

XX WO200197752-A2.  
 XX  
 XX 27-DEC-2001.  
 XX  
 XX 13-JUN-2001; 2001WO-US018976.  
 XX  
 XX 20-JUN-2000; 2000US-0212874P.  
 XX  
 XX (UYPE-) UNIV PENNSYLVANIA.  
 XX  
 XX Cines DB, Higazi AA;  
 XX  
 XX WPI; 2002-122240/16.  
 XX P-PSDB; AAE16542.  
 XX  
 XX Composition for modulating muscle cell and tissue contractility for  
 XX treating atherosclerosis, asthma, hypertension, glaucoma, impotence,  
 XX comprising domains from urokinase-type plasminogen activator.  
 XX  
 XX Claim 29; Fig 1J; 117pp; English.

XX The invention relates to a composition comprising one or more domains of  
 CC urokinase-type plasminogen activator (uPA). The composition is used to  
 CC modulate the contractility and angiogenic activity of a mammalian muscle,  
 CC endothelial cell or tissue. The composition is used for treating stroke,  
 CC hypotension, hypertension, atherosclerosis, heart attack, microvascular

occlusions, thrombotic microangiopathies, surgically induced thrombotic disorders, angiogenic disorders, pulmonary fibrosis, asthma, tumour cell invasion, tumour cell angiogenesis, tumour cell metastasis, glaucoma, diabetic retinopathy, wound healing, clotting disorder, uterine contraction disorder, male impotence, respiratory disease or condition such as asthma, adult respiratory distress syndrome, primary pulmonary hypertension, macrovascular thrombotic occlusion, and a disorder associated with chronic intrapulmonary fibrin formation. The present sequence is human urokinase-type plasminogen activator (uPA) kringle DNA

XX Sequence 264 BP; 66 A; 74 C; 74 G; 50 T; 0 U; 0 Other;

Query Match 72.3%; Score 18.8; DB 6; Length 264;

Best Local Similarity 90.9%; Pred. No. 1.3e-02; Indels 0; Gaps 0; Matches 20; Conservative 0; Mismatches 2;

QY 5 CTGGAAGTCTGCCAGTGTCTT 26  
|||||  
Db 78 CTGGAAGTCTGCCAGTGTCTT 99

## RESULT 12

AAD27083  
ID AAD27083 standard; DNA; 288 BP.

AC AAD27083;

DT 09-APR-2002 (first entry)

DE Human uPA kringle and connecting peptide DNA.

KW Human; urokinase-type plasminogen activator; uPA; therapy; hypertension; stroke; hypotension; atherosclerosis; heart attack; thrombotic disorder; microvascular occlusion; angiogenic disorder; pulmonary fibrosis; asthma; tumour cell metastasis; glaucoma; diabetic retinopathy; wound healing; clotting disorder; uterine contraction disorder; respiratory disease; adult respiratory distress syndrome; male impotence; ds.

XX Homo sapiens.

XX Key Location/Qualifiers

FT CDS 1..288

FT /tag= a

FT /product= "Human uPA kringle and connecting peptide"

FT /note= "CDS does not include start and stop codon"

FT /partial

XX WO200197752-A2.

XX PD 27-DEC-2001.

XX PF 13-JUN-2001; 2001WO-US018976.

XX PR 20-JUN-2000; 2000US-0212874P.

XX PA (UTYPE-) UNIV PENNSYLVANIA.

XX PI Cines DB, Higazi AA;

XX WPI; 2002-122240/16.

XX DR P-PSDB; AAB15550.

XX Composition for modulating muscle cell and tissue contractility for treating atherosclerosis, asthma, hypertension, glaucoma, impotence, comprising domains from urokinase-type plasminogen activator.

XX Claim 29; Fig 1R; 117pp; English.

XX The invention relates to a composition comprising one or more domains of urokinase-type plasminogen activator (uPA). The composition is used to modulate the contractility and angiogenic activity of a mammalian muscle, endothelial cell or tissue. The composition is used for treating stroke, hypotension, hypertension, atherosclerosis, heart attack, microvascular occlusions, thrombotic microangiopathies, surgically induced thrombotic disorders, angiogenic disorders, pulmonary fibrosis, asthma, tumour cell invasion, tumour cell angiogenesis, tumour cell metastasis, glaucoma, diabetic retinopathy, wound healing, clotting disorder, uterine contraction disorder, male impotence, respiratory disease or condition such as asthma, adult respiratory distress syndrome, primary pulmonary hypertension, macrovascular thrombotic occlusion, and a disorder associated with chronic intrapulmonary fibrin formation. The present sequence is human urokinase-type plasminogen activator (uPA) kringle DNA

XX occlusions, thrombotic microangiopathies, surgically induced thrombotic disorders, angiogenic disorders, pulmonary fibrosis, asthma, tumour cell invasion, tumour cell angiogenesis, tumour cell metastasis, glaucoma, diabetic retinopathy, wound healing, clotting disorder, uterine contraction disorder, male impotence, respiratory disease or condition such as asthma, adult respiratory distress syndrome, primary pulmonary hypertension, macrovascular thrombotic occlusion, and a disorder associated with chronic intrapulmonary fibrin formation. The present sequence is a DNA encoding human urokinase-type plasminogen activator (uPA) kringle and connecting peptide

XX Sequence 288 BP; 73 A; 84 C; 77 G; 54 T; 0 U; 0 Other;

Query Match 72.3%; Score 18.8; DB 6; Length 288;

Best Local Similarity 90.9%; Pred. No. 1.4e+02; Indels 0; Gaps 0; Matches 20; Conservative 0; Mismatches 2;

QY 5 CTGGAAGTCTGCCAGTGTCTT 26  
|||||  
Db 78 CTGGAAGTCTGCCAGTGTCTT 99

## RESULT 13

AAN81299  
ID AAN81299 standard; cDNA; 300 BP.

XX AAN81299;

XX AC

XX 25-MAR-2003 (revised)

DT 08-OCT-1990 (first entry)

XX Human pro-urokinase mutant designed to acquire unique restriction sites.

XX human pro-urokinase mutant; fibrinolysis; epidermal growth factor domain; ss.

XX Synthetic.

XX Key Location/Qualifiers

FT misc\_difference 69

FT /tag= a

FT /label= point mutation

FT /note= "Wild-type= A; Mutant= G"

FT misc\_difference 72

FT /tag= b

FT /label= point mutation

FT /note= "Wild-type= T; Mutant= C"

FT misc\_difference 209..210

FT /tag= c

FT /label= two point mutations

FT /note= "Wild-type= GC; Mutant= CA"

XX EP253241-A.

XX PD 20-JAN-1988.

XX PF 03-JUL-1987; 87EP-00109628.

XX PR 03-JUL-1986; 86JP-00156936.

XX DR 18-FEB-1987; 87JP-00036495.

XX (GRC) GREEN CROSS CORP.

XX Kasai S, Hiramatsu R, Uno S, Nagai M, Arimura H;

XX WPI; 1988-015623/03.

XX DR P-PSDB; AAP80996.

XX New human pro-urokinase mutants with fibrinolytic activity - have longer half-lives in blood and are obtained by recombinant DNA procedures.

XX Disclosure; Page ?; 40pp; English.



CC Site-directed mutagenesis creates two unique restriction sites; the first  
CC is recognised by SacI and the second by NdeI. The mutant sequence is  
CC inserted into an expression vector system for expression in eg E.coli.  
CC (Updated on 25-MAR-2003 to correct PR field.) (Updated on 25-MAR-2003 to  
CC correct PI field.)

XX SQ Sequence 300 BP; 77 A; 82 C; 75 G; 66 T; 0 U; 0 Other;  
Query Match 72.3%; Score 18.8; DB 1; Length 300;  
Best Local Similarity 90.9%; Pred. No. 1.4e+02;  
Matches 20; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 5 CTGGAAGCTCTGCCAGTCTCTT 26  
||||| ||||| ||||| ||||| |||||  
Db 279 CTGGAAGCTCTGCCAGTCTCTT 300

RESULT 14  
AAD27078  
ID AAD27078 standard; DNA; 405 BP.

XX AC AAD27078;

XX DT 09-APR-2002 (first entry)

XX DE Human urokinase-type plasminogen activator amino terminal fragment DNA.

XX KW Human; urokinase-type plasminogen activator; uPA; therapy; hypertension;  
XX stroke; hypotension; atherosclerosis; heart attack; thrombotic disorder;  
XX microvascular occlusion; angioecic disorder; pulmonary fibrosis; asthma;  
XX tumour cell metastasis; glaucoma; diabetic retinopathy; wound healing;  
XX clotting disorder; uterine contraction disorder; respiratory disease;  
XX adult respiratory distress syndrome; amino terminal fragment; ATF;  
XX male impotence; ds.

XX OS Homo sapiens.

XX FH Key Location/Qualifiers  
XX CDS 1..405

XX FT /\*tag= a  
XX FT /product= "Human uPA ATF"  
XX FT /note= "CDS does not include start and stop codon"  
XX FT /partial

XX FN WO200197752-A2.

XX PD 27-DEC-2001.

XX PF 13-JUN-2001; 2001WO-US018976.

XX PR 20-JUN-2000; 2000US-0212874P.

XX XX (UYPE-) UNIV PENNSYLVANIA.

XX PA Cines DB, Higazi AA;

XX PI WPI; 2002-122240/16.

XX DR P-PSDB; AAE16545.

XX XX Composition for modulating muscle cell and tissue contractility for  
XX PT treating atherosclerosis, asthma, hypertension, glaucoma, impotence,  
XX PT comprising domains from urokinase-type plasminogen activator.

XX PS Claim 29; Fig 1M; 117pp; English.

XX XX The invention relates to a composition comprising one or more domains of  
XX urokinase-type plasminogen activator (uPA). The composition is used to  
XX modulate the contractility and angiogenic activity of a mammalian muscle,  
XX endothelial cell or tissue. The composition is used for treating stroke,  
XX hypotension, hypertension, atherosclerosis, heart attack, microvascular  
XX occlusions, thrombotic microangiopathies, surgically induced thrombotic  
XX disorders, angiogenic disorders, pulmonary fibrosis, asthma, tumour cell  
XX invasion, tumour cell angiogenesis, tumour cell metastasis, glaucoma,

CC diabetic retinopathy, wound healing, clotting disorder, uterine  
CC contraction disorder, male impotence, respiratory disease or condition  
CC such as asthma, adult respiratory distress syndrome, primary pulmonary  
CC hypertension, microvascular thrombotic occlusion, and a disorder  
CC associated with chronic intrapulmonary fibrin formation. The present  
CC sequence is human urokinase-type plasminogen activator (uPA) amino  
XX terminal fragment (ATF) DNA

SQ Sequence 405 BP; 113 A; 105 C; 104 G; 83 T; 0 U; 0 Other;  
Query Match 72.3%; Score 18.8; DB 6; Length 405;  
Best Local Similarity 90.9%; Pred. No. 1.4e+02;  
Matches 20; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 5 CTGGAAGCTCTGCCAGTCTCTT 26  
||||| ||||| ||||| ||||| |||||  
Db 219 CTGGAAGCTCTGCCAGTCTCTT 240

RESULT 15

AAD27082

ID AAD27082 standard; DNA; 429 BP.

XX AC AAD27082;

XX DT 09-APR-2002 (first entry)

XX DE Human uPA amino terminal fragment (ATF) and connecting peptide DNA.

XX KW Human; urokinase-type plasminogen activator; uPA; therapy; hypertension;  
XX stroke; hypotension; atherosclerosis; heart attack; thrombotic disorder;  
XX microvascular occlusion; angioecic disorder; pulmonary fibrosis; asthma;  
XX tumour cell metastasis; glaucoma; diabetic retinopathy; wound healing;  
XX clotting disorder; uterine contraction disorder; respiratory disease;  
XX adult respiratory distress syndrome; amino terminal fragment; ATF;  
XX male impotence; ds.

XX OS Homo sapiens.

XX FH Key Location/Qualifiers  
XX CDS 1..429

XX FT /\*tag= a

XX FT /product= "Human uPA ATF and connecting peptide"

XX FT /note= "CDS does not include start and stop codon"  
XX FT /partial

XX FN WO200197752-A2.

XX PD 27-DEC-2001.

XX PF 13-JUN-2001; 2001WO-US018976.

XX PR 20-JUN-2000; 2000US-0212874P.

XX XX (UYPE-) UNIV PENNSYLVANIA.

XX PA Cines DB, Higazi AA;

XX PI WPI; 2002-122240/16.

XX DR P-PSDB; AAE16549.

XX XX Composition for modulating muscle cell and tissue contractility for  
XX PT treating atherosclerosis, asthma, hypertension, glaucoma, impotence,  
XX PT comprising domains from urokinase-type plasminogen activator.

XX PS Claim 29; Fig 1Q; 117pp; English.

XX XX The invention relates to a composition comprising one or more domains of  
XX urokinase-type plasminogen activator (uPA). The composition is used to  
XX modulate the contractility and angiogenic activity of a mammalian muscle,  
XX endothelial cell or tissue. The composition is used for treating stroke,  
XX hypotension, hypertension, atherosclerosis, heart attack, microvascular  
XX occlusions, thrombotic microangiopathies, surgically induced thrombotic

CC disorders, angiogenic disorders, pulmonary fibrosis, asthma, tumour cell  
 CC invasion, tumour cell angiogenesis, tumour cell metastasis, glaucoma,  
 CC diabetic retinopathy, wound healing, clotting disorder, uterine  
 CC contraction disorder, male impotence, respiratory disease or condition  
 CC such as asthma, adult respiratory distress syndrome, primary pulmonary  
 CC hypertension, microvascular thrombotic occlusion, and a disorder  
 CC associated with chronic intrapulmonary fibrin formation. The present  
 CC sequence is a DNA encoding human urokinase-type plasminogen activator  
 CC (uPA) amino terminal fragment (ATF) and connecting peptide  
 XX  
 SQ Sequence 429 BP; 120 A; 115 C; 107 G; 87 T; 0 U; 0 Other;

Query Match 72.3%; Score 18.8; DB 6; Length 429;  
 Best Local Similarity 90.9%; Pred. No. 1.4e-02;  
 Matches 20; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 5 CTGGAAGTCTGCCAGTGTCTT 26  
 |||||  
 Db 219 CTGGAACCTGCCACTGTCTT 240

Search completed: February 19, 2004, 21:51:43  
 Job time : 25.0879 secs

GenCore version 5.1.1.6  
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: February 19, 2004, 21:21:54 ; Search time 172.149 Seconds  
(without alignments)  
4510.152 Million cell updates/sec

Title: US-09-939-853a-141

Perfect score: 26

Sequence: 1 cctctggaagctgcagtgctcct 26

Scoring table: IDENTITY NUC

Gapop 10.0 , Gapext 1.0

Searched: 27513289 seqs, 14931090276 residues

Total number of hits satisfying chosen parameters: 55026578

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

```
EST:*
1: em_estba:*
2: em_esthum:*
3: em_estin:*
4: em_estmu:*
5: em_estov:*
6: em_estpl:*
7: em_estro:*
8: em_hic:*
9: gb_est1:*
10: gb_est2:*
11: gb_hic:*
12: gb_est3:*
13: gb_est4:*
14: gb_est5:*
15: em_estfun:*
16: em_estom:*
17: em_gss_hum:*
18: em_gss_inv:*
19: em_gss_pln:*
20: em_gss_vrt:*
21: em_gss_fun:*
22: em_gss_nam:*
23: em_gss_mus:*
24: em_gss_pro:*
25: em_gss_rod:*
26: em_gss_phg:*
27: em_gss_vrl:*
28: gb_gss1:*
29: gb_gss2:*
```

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Match	Length	ID	Description
C 1	26	100.0	616	13	BX383606
C 2	26	100.0	778	12	BG178487
C 3	26	100.0	878	12	BQ053486
C 4	26	100.0	986	12	BQ054265

C	5	26	100.0	1020	12	BQ054281
C 6	26	100.0	1201	9	AL541041	
C 7	21.8	83.8	354	12	BM483329	
C 8	21.8	83.8	389	28	AZ620995	
C 9	21.8	83.8	541	12	BI898999	
C 10	20.2	77.7	442	12	BG228158	
C 11	20.2	77.7	689	10	BB468570	
C 12	20.2	77.7	693	10	BB195256	
C 13	20.2	77.7	1226	11	AK011064	
C 14	20.2	77.7	3016	11	AK039889	
C 15	19.8	76.2	276	29	CG508158	
C 16	19.8	76.2	518	9	AA747812	
C 17	19.6	75.4	302	29	CG635108	
C 18	19.6	75.4	389	10	BE015329	
C 19	19.6	75.4	439	10	BE015326	
C 20	19.6	75.4	439	10	CA893892	
C 21	19.6	75.4	533	9	AI551288	
C 22	19.6	75.4	578	14	CA894739	
C 23	19.6	75.4	967	12	BI192570	
C 24	19.6	75.4	1114	28	CC250655	
C 25	19.6	75.4	1132	28	CC286280	
C 26	19.6	75.4	1281	28	CC283325	
C 27	19.4	74.6	546	13	C78860	
C 28	19.2	73.8	129	29	CG521539	
C 29	19.2	73.8	206	29	CG558956	
C 30	19.2	73.8	208	29	CG555648	
C 31	19.2	73.8	213	29	CG576620	
C 32	19.2	73.8	249	29	CG639953	
C 33	19.2	73.8	284	29	CG648011	
C 34	19.2	73.8	287	29	CG618235	
C 35	19.2	73.8	287	29	CG530122	
C 36	19.2	73.8	289	29	CG578637	
C 37	19.2	73.8	291	29	CG645423	
C 38	19.2	73.8	348	9	AI163788	
C 39	19.2	73.8	403	14	T74108	
C 40	19.2	73.8	438	14	T61390	
C 41	19.2	73.8	521	29	CE370901	
C 42	19.2	73.8	553	9	AA931151	
C 43	19.2	73.8	606	28	BH310531	
C 44	19.2	73.8	661	29	AG043855	
C 45	19.2	73.8	749	28	BZ204700	

#### ALIGNMENTS

RESULT 1  
BX383606/c  
LOCUS  
DEFINITION BX383606 Homo sapiens T CELLS (JURKAT CELL LINE) COT 10-NORMALIZED  
Homo sapiens cDNA clone CS0D013YK10 5-PRIME, mRNA sequence.  
ACCESSION BX383606  
VERSION BX383606.1 GI:30457152  
KEYWORDS EST.  
SOURCE Homo sapiens (human)  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo..  
1 (bases 1 to 616)  
Li, W.-S., Gruber, C., Jessee, J. and Polayes, D.  
Full-length cDNA libraries and normalization  
Unpublished (2001)  
JOURNAL  
COMMENT  
Genoscope - Centre National de Sequencage  
BP 191 91006 EVRY cedex - France  
Email: seqref@genoscope.cns.fr, Web : www.genoscope.cns.fr  
Library was constructed by Life Technologies, a division of  
Invitrogen. This sequence belongs to sequence cluster 9825.r. For  
more information about this cluster, see  
http://www.genoscope.cns.fr/  
cgi-bin/cluster.cgi?seq=CS0D013BF05QP1&cluster=9825.r. Contact :  
Feng Liang Email : fliang@lifetech.com URL :  
http://fulllength.invitrogen.com/ Invitrogen Corporation 1600

Faraday Avenue Genoscope sequence ID : CS0DJ013BF05QPI.

## FEATURES

source  
1. .616  
/organization="Homo sapiens"  
/mol\_type="mRNA"  
/db\_xref="taxon:9606"  
/clone="CS0DJ013YK10"  
/cell\_type="T CELLS (JURKAT CELL LINE) COT 10-NORMALIZED"  
/clone\_lib="Homo sapiens T CELLS (JURKAT CELL LINE) COT 10-NORMALIZED"  
/note="1st strand cDNA was primed with a NotI-oligo(dT) primer. Five prime end enriched, double-strand cDNA was digested with Not I and cloned into the Not I and EcoR V sites of the pCMVSPORT 6 vector. Library was normalized."

## ORIGIN

Query Match 100.0%; Score 26; DB 13; Length 616;  
Best Local Similarity 100.0%; Pred. No. 1.4;  
Matches 26; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 CCTCTGGAAGTCTGCCAGTGTCTT 26  
Db 325 CCTCTGGAAGTCTGCCAGTGTCTT 300

## RESULT 2

BG178487/c  
LOCUS  
DEFINITION 778 bp mRNA linear EST 06-FEB-2001  
60228305F1 NIH\_MGC\_91 Homo sapiens cDNA clone IMAGE:4429896 5',  
mRNA sequence.

ACCESSION BG178487

VERSION BG178487.1 GI:12685190

KEYWORDS EST.

SOURCE Homo sapiens (human)

## ORGANISM

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

## REFERENCE

1 (bases 1 to 778)  
NIH-MGC http://mgs.nci.nih.gov/.

National Institutes of Health, Mammalian Gene Collection (MGC)

Unpublished (1999)

Contact: Robert Strausberg, Ph.D.

Email: cgabbs@mail.nih.gov

Tissue Procurement: DCTD/DTF

cDNA Library Preparation: Life Technologies, Inc.

cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)

DNA Sequencing by: Incyte Genomics, Inc.

Clone distribution: MGC clone distribution information can be

found through the I.M.A.G.E. Consortium/LLNL at:

http://image.llnl.gov

Plate: L1AM10182 row: i column: 01

High quality sequence stop: 657.

Location/Qualifiers

## FEATURES

source  
1. .778  
/organization="Homo sapiens"  
/mol\_type="mRNA"  
/db\_xref="taxon:9606"  
/clone="IMAGE:4429896"  
/tissue\_type="adenocarcinoma, cell line"  
/lab\_host="DH10B (phage-resistant)"  
/clone\_lib="NIH MGC 91"  
/note="Organ: Prostate; Vector: pCMV-SPORT6; Site 1: NotI;  
Site 2: SalI; Cloned unidirectionally; oligo-dT primed.  
Average insert size 1.4 kb. Library enriched for  
full-length clones and constructed by Life Technologies.  
Note: this is a NIH\_MGC Library."

## ORIGIN

Query Match 100.0%; Score 26; DB 12; Length 778;  
Best Local Similarity 100.0%; Pred. No. 1.5;  
Matches 26; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 CCTCTGGAAGTCTGCCAGTGTCTT 26  
Db 240 CCTCTGGAAGTCTGCCAGTGTCTT 215

## RESULT 3

BQ053486/c

LOCUS

DEFINITION 878 bp mRNA linear EST 29-MAR-2002

AGENCOURT\_6822017 NIH\_MGC\_106 Homo sapiens cDNA clone IMAGE:5935253

5', mRNA sequence.

ACCESSION BQ053486

VERSION BQ053486.1 GI:19812826

KEYWORDS EST.

SOURCE Homo sapiens (human)

## ORGANISM

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

1 (bases 1 to 878)

NIH-MGC http://mgs.nci.nih.gov/.

National Institutes of Health, Mammalian Gene Collection (MGC)

Unpublished (1999)

Contact: Robert Strausberg, Ph.D.

Email: cgabbs@mail.nih.gov

Tissue Procurement: Dr. Daniel McVicar, PBS/NCI

cDNA Library Preparation: Rubin Laboratory

cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)

DNA Sequencing by: Agencourt Bioscience Corporation

Clone distribution: MGC clone distribution information can be

found through the I.M.A.G.E. Consortium/LLNL at:

http://image.llnl.gov

Plate: L1CM2122 row: i column: 06

High quality sequence stop: 394.

Location/Qualifiers

## FEATURES

source  
1. .878  
/organization="Homo sapiens"  
/mol\_type="mRNA"  
/db\_xref="taxon:9606"  
/clone="IMAGE:5935253"  
/tissue\_type="natural killer cells, cell line"  
/lab\_host="DH10B (phage-resistant)"  
/clone\_lib="NIH MGC 106"  
/note="Organ: blood; Vector: pOT87; Site 1: XhoI; Site 2:  
EcoRI; cDNA made by oligo-dT priming. Directionally cloned  
into EcoRI/XhoI sites using the following 5' adaptor:  
GGCAGCAG(G). Library constructed by Ling Hong in the  
Laboratory of Gerald M. Rubin (University of California,  
Berkeley) using ZAP-cDNA synthesis kit (Stratagene) and  
Superscript II RT (Life Technologies). Note: this is a  
NIH\_MGC Library."

## ORIGIN

Query Match 100.0%; Score 26; DB 12; Length 878;  
Best Local Similarity 100.0%; Pred. No. 1.5;  
Matches 26; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 CCTCTGGAAGTCTGCCAGTGTCTT 26  
Db 151 CCTCTGGAAGTCTGCCAGTGTCTT 126

## RESULT 4

BQ054265/c

LOCUS

DEFINITION 986 bp mRNA linear EST 29-MAR-2002

AGENCOURT\_6830248 NIH\_MGC\_106 Homo sapiens cDNA clone IMAGE:5936339

5', mRNA sequence.

ACCESSION BQ054265

VERSION BQ054265.1 GI:19813605

KEYWORDS EST.

SOURCE Homo sapiens (human)

## ORGANISM

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

1 (bases 1 to 986)

AUTHORS NIH-MGC <http://mgc.nci.nih.gov/>.  
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)  
JOURNAL Unpublished (1999)  
COMMENT Contact: Robert Strausberg, Ph.D.  
Email: cgabbs-remail.nih.gov  
Tissue Procurement: Dr. Daniel McVicar, DBS/NCI  
cDNA Library Preparation: Rubin Laboratory  
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)  
DNA Sequencing by: Agencourt Bioscience Corporation  
Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at: <http://image.llnl.gov>  
Plate: LLCM2125 row: i column: 12  
High quality sequence stop: 556.  
Location/Qualifiers  
1..986  
/organism="Homo sapiens"  
/mol\_type="mRNA"  
/db\_xref="taxon:9606"  
/clone="IMAGE:5936339"  
/tissue\_type="natural killer cells, cell line"  
/lab\_host="DH10B (phage-resistant)"  
/clone\_lib="NIH\_MGC\_106"  
/note="Organ: blood; Vector: pOTB7; Site: 1: XhoI; Site 2: EcoRI; cDNA made by oligo-dT priming. Directionally cloned into EcoRI/XhoI sites using the following 5' adaptor: GCACGAG(G). Library constructed by Ling Hong in the laboratory of Gerald M. Rubin (University of California, Berkeley) using ZAP-cDNA synthesis kit (Stratagene) and Superscript II RT (Life Technologies). Note: this is a NIH\_MGC Library."

ORIGIN  
Query Match 100.0%; Score 26; DB 12; Length 986;  
Best Local Similarity 100.0%; Pred. No. 1.6; Mismatches 0; Indels 0; Gaps 0;  
Matches 26; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 CCTTCTGGAAGTCTGCCAGTGTCTT 26  
|||||  
DB 157 CCTTCTGGAAGTCTGCCAGTGTCTT 132  
|||||

RESULT 5  
BQ054281/c  
LOCUS BQ054281 1020 bp mRNA linear EST 29-MAR-2002  
DEFINITION AGENCOURT\_6830234 NIH\_MGC\_106 Homo sapiens cDNA clone IMAGE:5936362  
5', mRNA sequence.  
ACCESSION BQ054281  
VERSION BQ054281.1 GI:19813621  
KEYWORDS EST.  
SOURCE Homo sapiens (human)  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
1 (bases 1 to 1020)  
Li, W.B., Gruber, C., Jessee, J. and Polayes, D.  
Full-length cDNA libraries and normalization  
Unpublished (2001)  
On Feb 15, 2001 this sequence version replaced gi:12871733.  
Contact: Genoscope  
Genoscope - Centre National de Sequencage  
BP 131 91006 EVRY cedex - France  
Email: seqref@genoscope.cns.fr, Web : [www.genoscope.cns.fr](http://www.genoscope.cns.fr)  
Library was constructed by Life Technologies, a division of Invitrogen. This sequence belongs to sequence cluster 9825.r For more information about this cluster, see <http://www.genoscope.cns.fr/cgi-bin/cluster.cgi?seq=CS0DE005AF12QP1&cluster=9825.r>. Contact : Feng Liang Email : [fliang@lifetech.com](mailto:fliang@lifetech.com) URL : <http://fulllength.invitrogen.com/Invitrogen> Corporation 1600 Paraday Avenue Genoscope sequence ID : CS0DE005AF12QP1.  
Location/Qualifiers  
1..1201  
/organism="Homo sapiens"  
/mol\_type="mRNA"  
/db\_xref="taxon:9606"  
/clone="CS0DE005YK23"  
/tissue\_type="PLACENTA"  
/clone\_lib="Homo sapiens PLACENTA"  
/note="Vector: pCMVSPORT 6; 1st strand cDNA was primed with a NotI-oligo(dT) primer. Five prime end enriched, double-strand cDNA was digested with Not I and cloned into the Not I and EcoRV sites of the pCMVSPORT 6 vector. Library was not normalized."

FEATURES  
source  
Query Match 100.0%; Score 26; DB 9; Length 1201;  
Best Local Similarity 100.0%; Pred. No. 1.7; Mismatches 0; Indels 0; Gaps 0;  
Matches 26; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 CCTTCTGGAAGTCTGCCAGTGTCTT 26  
|||||

/db\_xref="taxon:9606"  
/clone="IMAGE:5936362"  
/tissue\_type="natural killer cells, cell line"  
/lab\_host="DH10B (phage-resistant)"  
/clone\_lib="NIH\_MGC\_106"  
/note="Organ: blood; Vector: pOTB7; Site: 1: XhoI; Site 2: EcoRI; cDNA made by oligo-dT priming. Directionally cloned into EcoRI/XhoI sites using the following 5' adaptor: GCACGAG(G). Library constructed by Ling Hong in the laboratory of Gerald M. Rubin (University of California, Berkeley) using ZAP-cDNA synthesis kit (Stratagene) and Superscript II RT (Life Technologies). Note: this is a NIH\_MGC Library."

ORIGIN  
Query Match 100.0%; Score 26; DB 12; Length 1020;  
Best Local Similarity 100.0%; Pred. No. 1.6; Mismatches 0; Indels 0; Gaps 0;  
Matches 26; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 CCTTCTGGAAGTCTGCCAGTGTCTT 26  
|||||  
DB 90 CCTTCTGGAAGTCTGCCAGTGTCTT 65  
|||||

RESULT 6  
AL541041/c  
LOCUS AL541041 1201 bp mRNA linear EST 12-MAY-2003  
DEFINITION AL541041 Homo sapiens PLACENTA Homo sapiens cDNA clone CS0DE005YK23  
5-PRIME mRNA sequence.  
ACCESSION AL541041  
VERSION AL541041.2 GI:30544829  
KEYWORDS EST.  
SOURCE Homo sapiens (human)  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
1 (bases 1 to 1201)  
Li, W.B., Gruber, C., Jessee, J. and Polayes, D.  
Full-length cDNA libraries and normalization  
Unpublished (2001)  
On Feb 15, 2001 this sequence version replaced gi:12871733.  
Contact: Genoscope  
Genoscope - Centre National de Sequencage  
BP 131 91006 EVRY cedex - France  
Email: seqref@genoscope.cns.fr, Web : [www.genoscope.cns.fr](http://www.genoscope.cns.fr)  
Library was constructed by Life Technologies, a division of Invitrogen. This sequence belongs to sequence cluster 9825.r For more information about this cluster, see <http://www.genoscope.cns.fr/cgi-bin/cluster.cgi?seq=CS0DE005AF12QP1&cluster=9825.r>. Contact : Feng Liang Email : [fliang@lifetech.com](mailto:fliang@lifetech.com) URL : <http://fulllength.invitrogen.com/Invitrogen> Corporation 1600 Paraday Avenue Genoscope sequence ID : CS0DE005AF12QP1.  
Location/Qualifiers  
1..1201  
/organism="Homo sapiens"  
/mol\_type="mRNA"  
/db\_xref="taxon:9606"  
/clone="CS0DE005YK23"  
/tissue\_type="PLACENTA"  
/clone\_lib="Homo sapiens PLACENTA"  
/note="Vector: pCMVSPORT 6; 1st strand cDNA was primed with a NotI-oligo(dT) primer. Five prime end enriched, double-strand cDNA was digested with Not I and cloned into the Not I and EcoRV sites of the pCMVSPORT 6 vector. Library was not normalized."

FEATURES  
source  
Query Match 100.0%; Score 26; DB 9; Length 1201;  
Best Local Similarity 100.0%; Pred. No. 1.7; Mismatches 0; Indels 0; Gaps 0;  
Matches 26; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 CCTTCTGGAAGTCTGCCAGTGTCTT 26  
|||||



Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae; Bovidae; Bos.

#### REFERENCE

1 (bases 1 to 541)  
 Smith, T.P.L., Grosse, W.M., Preking, B.A., Roberts, A.J., Stone, R.T., Casas, E., Wray, J.E., White, J., Cho, J.J., Fahrenkrug, S.C., Bennett, G.L., Heaton, M.P., Laegreid, W.W., Rohrer, G.A., Chitko-Mckown, C.G., Perte, G., Holt, I., Karamycheva, S., Liang, F., Quackenbush, J. and Keele, J.W.  
 Sequence evaluation of four pooled-tissue normalized bovine cDNA libraries and construction of a gene index for cattle  
 Genome Res. 11 (4), 626-630 (2001)

#### TITLE

Genome Res. 11 (4), 626-630 (2001)

#### JOURNAL

1180013

#### MEDLINE

11282978

#### COMMENT

Contact: Smith TPL  
 USDA, ARS, US Meat Animal Research Center  
 PO Box 166, Clay Center, NE 68933-0166, USA  
 Tel: 402 762 4366  
 Fax: 402 762 4390  
 Email: smith@mail.marc.usda.gov  
 Single pass sequencing. Bases called and alt trimmed with phred  
 v0.980904.e. Vector identified by cross\_match with the -minscore 18  
 and -mismatch 12 options.  
 PCR Primers

FORWARD: AGGAACAGCTATGACCAT

BACKWARD: GTTTTCCCGAGTCACGACG

Plate: 105 row: P column: 20

Sq primer: ATTAGTGACACTATAG.

#### FEATURES

source

Location/Qualifiers  
 1..541  
 /organism="Bos taurus"  
 /mol\_type="mRNA"  
 /db\_xref="taxon:9913"  
 /tissue\_type="pooled"  
 /lab\_host="DH108"  
 /clone\_lib="MARC 280V"  
 /notes="Vector: pCMV SPORT6; Site 1: NotI; Site 2: SalI;  
 Library made from pooled tissue from testis, thymus,  
 semitendinosus muscle, longissimus muscle, pancreas,  
 adrenal, and endometrium."

#### RESULT 11

BG228158/c

#### LOCUS

BG228158

#### DEFINITION

musculus

musculus

musculus

musculus

musculus

musculus

musculus

musculus

musculus

musculus

musculus

musculus

musculus

musculus

musculus

musculus

musculus

musculus

musculus

musculus

musculus

musculus

musculus

musculus

musculus

musculus

musculus

musculus

musculus

musculus

musculus

musculus

musculus

musculus

musculus

musculus

musculus

High quality sequence stop: 402.

Location/Qualifiers

source

1..442

/organism="Mus musculus"

/mol\_type="mRNA"

/strain="C57BL/6J"

/db\_xref="taxon:10090"

/clone="IMAGE:4001772"

/sex="male"

/tissue\_type="Spleen"

/dev\_stage="4 weeks"

/lab\_host="DH108"

/clone\_lib="Soares mouse 3NbMS"

/notes="Vector: pRTD3-pac (Pharmacia) with a modified

polylinker; Site 1: Not I; Site 2: Eco RI; 1st strand cDNA

was primed with a Not I - oligo(dT) primer [5'

TGTTACCAATCTGAAGTGGAGCGCGCGCTGTTTTTTTTTTTTTTTT

3']; double-stranded cDNA was ligated to Eco RI adaptors

(Pharmacia), digested with Not I and cloned into the Not I

and Eco RI sites of the modified pT73 vector. RNA

provided by Dr. Bertrand Jordan. Library went through

three rounds of normalization, and was constructed by

Bento Soares and M. Fatima Bonaldo."

Query Match 77.7%; Score 20.2; DB 12; Length 442;

Best Local Similarity 88.0%; Pred. No. 3.8e+02;

Matches 22; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 2 CTTCTGGAAGTCTGCCAGTGTCTT 26

||||| ||||| ||||| ||||| |||||

DB 232 CTTCTGGAAGTCTGCCAGTGTCTT 208

#### RESULT 11

BG228158/c

#### LOCUS

BG228158

#### DEFINITION

musculus

musculus

musculus

musculus

musculus

musculus

musculus

musculus

musculus

musculus

musculus

musculus

musculus

musculus

musculus

musculus

musculus

musculus

musculus

musculus

musculus

musculus

musculus

musculus

musculus

musculus

musculus

musculus

musculus

musculus

musculus

musculus

musculus

musculus

musculus

musculus

musculus

musculus

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae; Bovidae; Bos.

#### REFERENCE

1 (bases 1 to 541)  
 Smith, T.P.L., Grosse, W.M., Preking, B.A., Roberts, A.J., Stone, R.T., Casas, E., Wray, J.E., White, J., Cho, J.J., Fahrenkrug, S.C., Bennett, G.L., Heaton, M.P., Laegreid, W.W., Rohrer, G.A., Chitko-Mckown, C.G., Perte, G., Holt, I., Karamycheva, S., Liang, F., Quackenbush, J. and Keele, J.W.  
 Sequence evaluation of four pooled-tissue normalized bovine cDNA libraries and construction of a gene index for cattle  
 Genome Res. 11 (4), 626-630 (2001)

#### TITLE

Genome Res. 11 (4), 626-630 (2001)

#### JOURNAL

1180013

#### MEDLINE

11282978

#### COMMENT

Contact: Smith TPL  
 USDA, ARS, US Meat Animal Research Center  
 PO Box 166, Clay Center, NE 68933-0166, USA  
 Tel: 402 762 4366  
 Fax: 402 762 4390  
 Email: smith@mail.marc.usda.gov  
 Single pass sequencing. Bases called and alt trimmed with phred  
 v0.980904.e. Vector identified by cross\_match with the -minscore 18  
 and -mismatch 12 options.  
 PCR Primers

FORWARD: AGGAACAGCTATGACCAT

BACKWARD: GTTTTCCCGAGTCACGACG

Plate: 105 row: P column: 20

Sq primer: ATTAGTGACACTATAG.

#### FEATURES

source

Location/Qualifiers  
 1..541  
 /organism="Bos taurus"  
 /mol\_type="mRNA"  
 /db\_xref="taxon:9913"  
 /tissue\_type="pooled"  
 /lab\_host="DH108"  
 /clone\_lib="MARC 280V"  
 /notes="Vector: pCMV SPORT6; Site 1: NotI; Site 2: SalI;  
 Library made from pooled tissue from testis, thymus,  
 semitendinosus muscle, longissimus muscle, pancreas,  
 adrenal, and endometrium."

#### RESULT 11

BG228158/c

#### LOCUS

BG228158

#### DEFINITION

musculus

musculus

musculus

musculus

musculus

musculus

musculus

musculus

musculus

musculus

musculus

musculus

musculus

musculus

musculus

musculus

musculus

musculus

musculus

musculus

musculus

musculus

musculus

musculus

musculus

musculus

musculus

musculus

musculus

musculus

musculus

musculus

musculus

musculus

musculus

musculus

musculus

musculus

High quality sequence stop: 402.

Location/Qualifiers

source

1..442

/organism="Mus musculus"

/mol\_type="mRNA"

/strain="C57BL/6J"

/db\_xref="taxon:10090"

/clone="IMAGE:4001772"

/sex="male"

/tissue\_type="Spleen"

/dev\_stage="4 weeks"

/lab\_host="DH108"

/clone\_lib="Soares mouse 3NbMS"

/notes="Vector: pRTD3-pac (Pharmacia) with a modified

polylinker; Site 1: Not I; Site 2: Eco RI; 1st strand cDNA

was primed with a Not I - oligo(dT) primer [5'

TGTTACCAATCTGAAGTGGAGCGCGCGCTGTTTTTTTTTTTTTTTT

3']; double-stranded cDNA was ligated to Eco RI adaptors

(Pharmacia), digested with Not I and cloned into the Not I

and Eco RI sites of the modified pT73 vector. RNA

provided by Dr. Bertrand Jordan. Library went through

three rounds of normalization, and was constructed by

Bento Soares and M. Fatima Bonaldo."

Query Match 77.7%; Score 20.2; DB 12; Length 442;

Best Local Similarity 88.0%; Pred. No. 3.8e+02;

Matches 22; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 2 CTTCTGGAAGTCTGCCAGTGTCTT 26

||||| ||||| ||||| ||||| |||||

DB 232 CTTCTGGAAGTCTGCCAGTGTCTT 208

#### RESULT 11

BG228158/c

#### LOCUS

BG228158

#### DEFINITION

musculus

musculus

musculus

musculus

musculus

musculus

musculus

musculus

musculus

musculus

musculus

musculus

musculus

musculus

musculus

musculus

musculus

musculus

musculus

musculus

musculus

musculus

musculus

musculus

musculus

musculus

musculus

musculus

musculus

musculus

musculus

musculus





3'. cDNA was cleaved with XhoI and BamHI. Vector: a modified pBluescript KS(+) after bulk excision from Lambda FLC I."

ORIGIN

Query Match 77.7%; Score 20.2; DB 10; Length 693;  
 Best Local Similarity 88.0%; Pred. No. 4.5e+02;  
 Matches 22; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 2 CTTCTGGAAGTCTGCAGTGTCTT 26  
 ||||| ||||| ||||| ||||| |||||  
 Db 595 CTTCTGGAAGTCTGCAGTGTCTT 571

RESULT 13  
 AK011064  
 LOCUS  
 DEFINITION Mus musculus 13 days embryo liver cDNA, RIKEN full-length enriched library, clone:2510040D07 product:hypothetical signal  
 TITLE peptidase/Trp-Asp (WD) repeats profile/Trp-Asp (WD) repeats  
 KEYWORDS circular profile/G-protein beta WD-40 repeats containing protein, full insert sequence.  
 ACCESSION AK011064  
 VERSION AK011064.1 GI:12846940  
 KEYWORDS HTC; CAP trapper.  
 SOURCE Mus musculus (house mouse)  
 ORGANISM Mus musculus  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

REFERENCE 1  
 Carninci, P. and Hayashizaki, Y.  
 High-efficiency full-length cDNA cloning  
 Meth. Enzymol. 303, 19-44 (1999)  
 JOURNAL 99279253  
 MEDLINE 10349636  
 PUBMED

REFERENCE 2  
 Carninci, P., Shibata, Y., Hayatsu, N., Sugahara, Y., Shibata, K., Itoh, M., Konno, H., Okazaki, Y., Muramatsu, M. and Hayashizaki, Y.  
 Normalization and subtraction of cap-trapper-selected cDNAs to prepare full-length cDNA libraries for rapid discovery of new genes  
 Genome Res. 10 (10), 1617-1630 (2000)  
 JOURNAL 20499374  
 MEDLINE 11042159  
 PUBMED

REFERENCE 3  
 Shibata, K., Itoh, M., Aizawa, K., Nagaoka, S., Sasaki, N., Carninci, P., Konno, H., Akiyama, J., Nishi, K., Kitsumai, T., Tashiro, H., Itoh, M., Sumi, N., Ishii, Y., Nakamura, S., Hazama, M., Nishine, T., Harada, A., Yamamoto, R., Matsumoto, H., Sakaguchi, S., Ikegami, T., Kashiwagi, K., Fujiwaka, S., Inoue, K., Togawa, Y., Izawa, M., Ohara, E., Watahiki, M., Yoneda, Y., Ishikawa, T., Ozawa, K., Tanaka, T., Matsuura, S., Kawai, J., Okazaki, Y., Muramatsu, M., Inoue, Y., Kira, A. and Hayashizaki, Y.  
 RIKEN integrated sequence analysis (RISA) system-384-format  
 sequencing pipeline with 384 multicapillary sequencer  
 Genome Res. 10 (11), 1757-1771 (2000)  
 JOURNAL 20530913  
 MEDLINE 11076861  
 PUBMED

REFERENCE 4  
 The RIKEN Genome Exploration Research Group Phase II Team and the FANTOM Consortium.  
 Functional annotation of a full-length mouse cDNA collection  
 Nature 409, 585-590 (2001)

REFERENCE 5  
 The FANTOM Consortium and the RIKEN Genome Exploration Research Group Phase I & II Team.  
 Analysis of the mouse transcriptome based on functional annotation of 60,770 full-length cDNAs  
 Nature 420, 563-573 (2002)

REFERENCE 6  
 (bases 1 to 1226)  
 Adachi, J., Aizawa, K., Akahira, S., Akimura, T., Arai, A., Aono, H., Arakawa, T., Bono, H., Carninci, P., Fukuda, S., Fukunishi, Y., Furuno, M., Hanagaki, T., Hara, A., Hayatsu, N., Hiramoto, K., Hiraka, T., Hori, F., Imotani, K., Ishii, Y., Itoh, M., Izawa, M., Kasukawa, T., Kato, H., Kawai, J., Kojima, Y., Konno, H., Kouda, M.,

Koya, S., Kurihara, C., Matsuyama, T., Miyazaki, A., Nishi, K., Nomura, K., Numazaki, R., Ohno, M., Okazaki, Y., Okido, T., Owa, C., Saito, H., Saito, R., Sakai, C., Sakai, K., Sano, H., Sasaki, D., Shibata, K., Shibata, Y., Shinagawa, A., Shiraki, T., Sogabe, Y., Suzuki, H., Tagami, M., Tagawa, A., Takahashi, F., Tanaka, T., Tejima, Y., Toya, T., Yamamura, T., Yasunishi, A., Yoshida, K., Yoshino, M., Muramatsu, M. and Hayashizaki, Y.  
 Direct Submission  
 Submitted (10-JUL-2000) Yoshihide Hayashizaki, The Institute of Physical and Chemical Research (RIKEN), Laboratory for Genome Exploration Research Group, RIKEN Genomic Sciences Center (GSC), RIKEN Yokohama Institute; 1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan (E-mail: genome-res@gsc.riken.go.jp, URL: http://genome.gsc.riken.go.jp/, Tel: 81-45-503-9222, Fax: 81-45-503-9216)  
 Please visit our web site (http://genome.gsc.riken.go.jp/) for further details.  
 cDNA library was prepared and sequenced in Mouse Genome Encyclopedia Project of Genome Exploration Research Group in Riken Genomic Sciences Center and Genome Science Laboratory in RIKEN Division of Experimental Animal Research in Riken contributed to prepare mouse tissues. First strand cDNA was primed with a primer [5' GAGAGAGAGAGATCCAGAGCTCTTTTITTTTNN 3'], cDNA was prepared by using trehalose thermo-activated reverse transcriptase and subsequently enriched for full-length by cap-trapper. Second strand cDNA was prepared with the primer adapter of sequence [5' GAGAGAGATTCGTGAGTTAATAATTAATCCGCCCCCCCCC 3']. cDNA was cleaved with XhoI and SstI. Cloning sites, 5' end: XhoI; 3' end: SstI.  
 Host: SOUR.

FEATURES

Location/Qualifiers

1..1226  
 /organism="Mus musculus"  
 /mol\_type="mRNA"  
 /strain="C57BL/6J"  
 /db\_xref="FANTOM DB:2510040D07"  
 /db\_xref="XGI:1902545"  
 /clone="2510040D07"  
 /tissue\_type="liver"  
 /clone\_lib="RIKEN full-length enriched mouse cDNA library"  
 /dev\_stage="13 days embryo"  
 60..1130  
 /note="unnamed protein product; hypothetical signal peptidase/Trp-Asp (WD) repeats profile/Trp-Asp (WD) repeats circular profile/G-protein beta WD-40 repeats containing protein (InterPro|IPR00508, PROSITE|PS50082, PROSITE|PS50294, InterPro|IPR001680, evidence: InterPro|putative"  
 /codon\_start=1  
 /protein\_id="BAB27371.1"  
 /db\_xref="GI:12846941"  
 /translation="MAAPSDPSLERHFKHDAVTRVDFSLNFKHLAGSMDSTLMI WHMKPQSRAYRFTGHNDVATCVNFPSPGHLASGSDKTVRIWPNVKGESTVFRHATVRSVHFCSDGSLVTSADKTVKWSRTHRQFLSLTQHINWCAKPSDGRLLIV SASDDKTVLWMDTRECISYCEHGFTVYDFHPSGTCIAAGMDTVKWDARTV RLLQHVOLHSAANVALSFHPSGNLYITAGSDSTLKILDMEGRLLYTLHGQGPATV AFRTGEVFASGGSDQVMVKSNFDIVYGDMAKRRPPLTSSSTLTIVSILEQRILTDRLLKQCLENQQLIMQRTTP"  
 1208..1213  
 /note="putative"  
 1226  
 /note="putative"

polyA\_signal  
 polyA\_site

ORIGIN

Query Match 77.7%; Score 20.2; DB 11; Length 1226;  
 Best Local Similarity 88.0%; Pred. No. 5.5e+02;  
 Matches 22; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 2 CTTCTGGAAGTCTGCAGTGTCTT 26  
 ||||| ||||| ||||| ||||| |||||  
 Db 1018 CTTCTGGAAGTCTGCAGTGTCTT 1042

RESULT 14  
 AK039689/c  
 LOCUS  
 DEFINITION  
 Mus musculus adult male spinal cord cDNA, RIKEN full-length enriched library, clone:A330087124 product:unclassifiable, full insert sequence.  
 AK039689 3016 bp mRNA linear HTC 19-SEP-2003  
 AK039689.1 GI:26087341  
 VERSION  
 HTc; CAP trapper.  
 KEYWORDS  
 Mus musculus (house mouse)  
 ORGANISM  
 Mus musculus  
 Eukaryota; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
 1  
 Carninci, P. and Hayashizaki, Y.  
 TITLE  
 High-efficiency full-length cDNA cloning  
 JOURNAL  
 Meth. Enzymol. 303, 19-44 (1999)  
 MEDLINE  
 99279253  
 PUBMED  
 10349636  
 2  
 Carninci, P., Shibata, Y., Hayatsu, N., Sugahara, Y., Shibata, K., Itoh, M., Konno, H., Okazaki, Y., Muramatsu, M. and Hayashizaki, Y.  
 TITLE  
 Normalization and subtraction of cap-trapper-selected cDNAs to prepare full-length cDNA libraries for rapid discovery of new genes  
 JOURNAL  
 Genome Res. 10 (10), 1617-1630 (2000)  
 MEDLINE  
 20499374  
 PUBMED  
 11042159  
 3  
 Shibata, K., Itoh, M., Aizawa, K., Nagasaka, S., Sasaki, N., Carninci, P., Konno, H., Akiyama, J., Nishi, K., Kitsuana, T., Tashiro, H., Itoh, M., Sumi, N., Ishii, Y., Nakamura, S., Hazama, M., Nishine, T., Harada, A., Yanamoto, R., Matsumoto, H., Sakaguchi, S., Ikegami, T., Kashiwagi, K., Fujiwaki, S., Inoue, K., Togawa, Y., Izawa, M., Ohara, E., Watahiki, M., Yoneda, Y., Ishikawa, T., Ozawa, K., Tanaka, T., Matsunura, S., Kawai, J., Okazaki, Y., Muramatsu, M., Inoue, Y., Kira, A. and Hayashizaki, Y.  
 TITLE  
 RIKEN integrated sequence analysis (RISA) system--384-format sequencing pipeline with 384 multipipillary sequencer  
 JOURNAL  
 Genome Res. 10 (11), 1757-1771 (2000)  
 MEDLINE  
 20530913  
 PUBMED  
 11076861  
 4  
 The RIKEN Genome Exploration Research Group Phase II Team and the FANTOM Consortium.  
 TITLE  
 Functional annotation of a full-length mouse cDNA collection  
 JOURNAL  
 Nature 409, 685-690 (2001)  
 MEDLINE  
 11076861  
 5  
 The FANTOM Consortium and the RIKEN Genome Exploration Research Group Phase I & II Team.  
 TITLE  
 Analysis of the mouse transcriptome based on functional annotation of 60,770 full-length cDNAs  
 JOURNAL  
 Nature 420, 563-573 (2002)  
 MEDLINE  
 12000000  
 PUBMED  
 12000000  
 6  
 (bases 1 to 3016)  
 Adachi, J., Aizawa, K., Akimura, T., Arakawa, T., Bono, H., Carninci, P., Fukuda, S., Furuno, M., Hanagaki, T., Hara, A., Hashizume, M., Hayashida, K., Hayatsu, N., Hiramoto, K., Hiraoka, T., Hirozane, T., Hori, F., Imotani, K., Ishii, Y., Itoh, M., Kagawa, I., Kasukawa, T., Katoh, H., Kawai, J., Kojima, Y., Kondo, S., Konno, H., Kouda, M., Koyu, S., Kurikara, C., Matsuyama, T., Miyazaki, A., Murata, M., Nakamura, M., Nishi, K., Nomura, K., Numazaki, R., Ohno, M., Ohsato, N., Okazaki, Y., Saito, R., Saitoh, H., Sakai, C., Sakai, K., Sakazume, N., Sano, H., Sasaki, D., Shibata, K., Shinagawa, A., Shiraki, T., Sogabe, Y., Tagami, M., Tagawa, A., Takahashi, F., Takaku-Akahira, S., Takeda, Y., Tanaka, T., Tomaru, A., Toya, T., Yasunishi, A., Muramatsu, M. and Hayashizaki, Y.  
 TITLE  
 Direct Submission  
 JOURNAL  
 Submitted (16-JUL-2001) Yoshihide Hayashizaki, The Institute of Physical and Chemical Research (RIKEN), Laboratory for Genome Exploration Research Group, RIKEN Genomic Sciences Center (GSC), RIKEN Yokohama Institute, 1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan (E-mail: genome-res@gsc.riken.go.jp).  
 URL: http://genome.gsc.riken.go.jp/ Tel: 81-45-503-9222, Fax: 81-45-503-9216  
 COMMENT  
 cDNA library was prepared and sequenced in Mouse Genome

Encyclopedia Project of Genome Exploration Research Group in Riken Genomic Sciences Center and Genome Science Laboratory in RIKEN. Division of Experimental Animal Research in Riken contributed to prepare mouse tissues.  
 Please visit our web site for further details.  
 URL: http://genome.gsc.riken.go.jp/  
 URL: http://fantom.gsc.riken.go.jp/.

FEATURES  
 source  
 location/Qualifiers  
 1..3016  
 /organism="Mus musculus"  
 /mol\_type="mRNA"  
 /strain="CS7BL/6J"  
 /db\_xref="PANTOM\_DB:A330087124"  
 /db\_xref="MGI:2404195"  
 /db\_xref="taxon:10090"  
 /clone="A330087124"  
 /sex="male"  
 /tissue\_type="spinal cord"  
 /clone\_lib="RIKEN full-length enriched mouse cDNA library"  
 /dev\_stage="adult"  
 1..3016  
 /note="unclassifiable"  
 misc\_feature

ORIGIN  
 Query Match 77.7%; Score 20.2; DB 11; Length 3016;  
 Best Local Similarity 88.0%; Pred. No. 7.6e+02;  
 Matches 22; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 2 CTTCTGGAAGTCTGCCAGTCTCCTT 26  
 Db 2918 CTTCTGGAAGTCTGCCAGTCTCCTT 2894

RESULT 15  
 CG508158/c  
 LOCUS  
 CG508158 276 bp DNA linear GSS 01-OCT-2003  
 OST58614 Mus musculus 129Sv/Ev Mus musculus genomic clone OST58614,  
 genomic survey sequence.  
 CG508158  
 CG508158.1 GI:37290088  
 VERSION  
 GSS.  
 KEYWORDS  
 Mus musculus (house mouse)  
 ORGANISM  
 Mus musculus  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
 1 (bases 1 to 276)  
 Zambrowicz, B.P., Abuin, A., Ramirez-Solis, R., Richter, L.J., Piggett, J., BeltrandelRio, H., Buxton, E.C., Edwards, J., Finch, R.A., Fiddle, C.J., Gupta, A., Hansen, G., Hu, Y., Huang, W., Jaing, C., Key, B.W. Jr., Kipp, P., Kohlhauff, B., Ma, Z.-Q., Markesich, D., Payne, R., Potter, D.G., Qian, N., Shaw, J., Schrick, J., Shi, Z.-Z., Sparks, M.J., Van Sligtenhorst, I., Vogel, P., Walke, W., Xu, N., Zhu, Q., Person, C. and Sands, A.I.  
 TITLE  
 Wnk1 kinase deficiency lowers blood pressure in mice: a gene-trap screen to identify potential targets for therapeutic intervention  
 JOURNAL  
 Proc. Natl. Acad. Sci. U.S.A. 100 (24), 14109-14114 (2003)  
 COMMENT  
 Contact: Zambrowicz Bp  
 OmniBank  
 Lexicon Genetics Incorporated  
 4000 Research Forest Drive, The Woodlands, TX 77381, USA  
 Email: materials@lexgen.com  
 Gene trap sequence tag generated by 3' RACE from mouse ES cells as described in Zambrowicz et al (Nature. 1998 Apr 9;392(6676):608-11)  
 Class: Gene Trap.

FEATURES  
 source  
 location/Qualifiers  
 1..276  
 /organism="Mus musculus"  
 /mol\_type="genomic DNA"  
 /strain="129Sv/Ev"  
 /db\_xref="taxon:10090"  
 /clone="OST58614"  
 /cell\_type="embryonic stem cell"  
 /clone\_lib="Mus musculus 129Sv/Ev"

ORIGIN

Query Match 76.2%; Score 19.8; DB 29; Length 276;  
 Best Local Similarity 84.0%; Pred. No. 4.8e+02;  
 Matches 21; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

Qy 2 CTTCTGGAGTCTGCCAGTGTCTT 26  
 |||||  
 Db 183 CTTCTGGACTTGTGCAATGTCTT 159  
 |||||

Search completed: February 20, 2004, 01:41:40  
 Job time : 176.149 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: February 19, 2004, 21:32:00 ; Search time 4.21902 Seconds  
(without alignments)  
3419.919 Million cell updates/sec

Title: US-09-939-853A-141  
Perfect score: 26  
Sequence: 1 cctctggaagtcgccagtcctt 26

Scoring table: IDENTITY\_NUC  
Gapop 10.0 , Gapext 1.0

Searched: 682709 seqs, 277475446 residues

Total number of hits satisfying chosen parameters: 1365418

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : Issued Patents NA: \*  
1: /cgn2\_6/prodata/2/ina/5A-COMB.seq: \*  
2: /cgn2\_6/prodata/2/ina/5B-COMB.seq: \*  
3: /cgn2\_6/prodata/2/ina/6A-COMB.seq: \*  
4: /cgn2\_6/prodata/2/ina/6B-COMB.seq: \*  
5: /cgn2\_6/prodata/2/ina/PCUS-COMB.seq: \*  
6: /cgn2\_6/prodata/2/ina/backfiles1.seq: \*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match %	Length	ID	Description
1	18.8	72.3	423	2	US-08-797-689-11
2	18.8	72.3	423	4	US-09-984-186-11
3	18.8	72.3	600	4	US-09-101-272G-72
4	18.8	72.3	624	4	US-09-101-272G-79
5	18.8	72.3	645	4	US-09-101-272G-95
6	18.8	72.3	666	4	US-09-101-272G-97
7	18.8	72.3	1233	1	US-08-254-922-1
8	18.8	72.3	1233	1	US-08-286-748B-1
9	18.8	72.3	1236	1	US-07-957-039A-7
10	18.8	72.3	1236	1	US-08-153-799-17
11	18.8	72.3	1236	4	US-09-023-655-927
12	18.8	72.3	1372	6	5219569-1
13	18.8	72.3	1475	4	US-09-643-597-122
14	18.8	72.3	1475	4	US-09-480-884A-122
15	18.8	72.3	1475	4	US-09-542-615A-122
16	18.8	72.3	1475	4	US-09-606-421B-122
17	18.8	72.3	1475	4	US-09-221-107-123
18	18.8	72.3	2294	4	US-09-643-597-123
19	18.8	72.3	2294	4	US-09-480-884A-123
20	18.8	72.3	2294	4	US-09-542-615A-123
21	18.8	72.3	2294	4	US-09-606-421B-123
22	18.8	72.3	2294	4	US-09-023-655-1217
23	18.8	72.3	2294	4	US-09-221-107-123
24	18.8	72.3	2301	6	5188829-2
25	17.8	68.5	9351	4	US-09-562-702A-11
26	17.8	68.5	9511	4	US-09-562-702A-9
27	17.6	67.7	273	4	US-09-313-294A-38

c

## ALIGNMENTS

### RESULT 1

US-08-797-689-11  
; Sequence 11, Application US/08797689  
; Patent No. 5876969  
; GENERAL INFORMATION:  
; APPLICANT: Fleer, Reinhard  
; APPLICANT: Fournier, Alain  
; APPLICANT: Guitton, Jean-Dominique  
; APPLICANT: Jeh, Gerard  
; APPLICANT: Yeh, Patrice  
; TITLE OF INVENTION: NOVEL BIOLOGICALLY ACTIVE POLYPEPTIDES,  
; TITLE OF INVENTION: PREPARATION THEREOF AND PHARMACEUTICAL COMPOSITION  
; TITLE OF INVENTION: CONTAINING SAID POLYPEPTIDES  
; NUMBER OF SEQUENCES: 36  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Rhone-Poulenc Rorer Inc.  
; STREET: 500 Arcola Road, 3C43  
; CITY: Collegeville  
; STATE: PA  
; COUNTRY: USA  
; ZIP: 19426  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: Macintosh  
; OPERATING SYSTEM: System 7.1  
; SOFTWARE: Word 5.1 (patentIn)  
; CURRENT APPLICATION DATA: US/08797,689  
; APPLICATION NUMBER: US/08797,689  
; FILING DATE: 31-JAN-1997  
; CLASSIFICATION: 435  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 08/256,927  
; FILING DATE: 28-JUL-1994  
; APPLICATION NUMBER: FR 92/01064  
; FILING DATE: 31-JAN-1992  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: PCT/FR93/00085  
; FILING DATE: 28-JAN-1993  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Smith Ph.D., Julie K.  
; REGISTRATION NUMBER: P-38,619  
; REFERENCE/DOCKET NUMBER: ST92006-US  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (610) 454-3839  
; TELEFAX: (610) 454-3808  
; INFORMATION FOR SEQ ID NO: 11:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 423 base pairs  
; TYPE: nucleic acid  
; STRANDEDNESS: double

c	28	17.6	67.7	908	4	US-09-800-729-25	Sequence 25, Appl
c	29	17.2	66.2	1335	1	US-07-942-157A-2	Sequence 2, Appl
c	30	17	65.4	274	4	US-09-313-294A-3335	Sequence 3335, Ap
c	31	17	65.4	449	4	US-09-621-976-14601	Sequence 14601, A
c	32	17	65.4	501	4	US-09-621-976-1584	Sequence 1584, Ap
c	33	17	65.4	1001	4	US-09-641-638-285	Sequence 285, App
c	34	17	65.4	3364	2	US-08-735-609-9	Sequence 9, Appl
c	35	17	65.4	3364	2	US-08-735-609-9	Sequence 9, Appl
c	36	17	65.4	3364	3	US-09-315-372-9	Sequence 9, Appl
c	37	17	65.4	3364	3	US-09-244-752-9	Sequence 9, Appl
c	38	17	65.4	3364	3	US-09-245-497-9	Sequence 9, Appl
c	39	17	65.4	3364	4	US-09-562-919-9	Sequence 9, Appl
c	40	17	65.4	8147	4	US-09-514-247A-9	Sequence 9, Appl
c	41	16.6	63.8	2157	1	US-08-336-618-25	Sequence 25, Appl
c	42	16.6	63.8	2200	2	US-08-462-481-3	Sequence 3, Appl
c	43	16.6	63.8	2200	2	US-08-436-771-5	Sequence 5, Appl
c	44	16.6	63.8	2200	2	US-08-434-998-5	Sequence 5, Appl
c	45	16.6	63.8	2200	2	US-08-487-797-5	Sequence 5, Appl

```

; TOPOLOGY: linear
; MOLECULE TYPE: cDNA
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 3..419
; US-08-797-689-11
Query Match 72.3%; Score 18.8; DB 2; Length 423;
Best Local Similarity 90.9%; Pred. No. 11;
Matches 20; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

OY 5 CTGGAAGTCTGCCAGTGTCTT 26
Db 230 CTGGAAGTCTGCCAGTGTCTT 251

RESULT 2
US-09-984-186-11
; Sequence 11, Application US/09984186
; Patent No. 6686179
; GENERAL INFORMATION:
; APPLICANT: Fleer, Reinhard
; Fournier, Alain
; Guitton, Jean-Dominique
; Jung, Gerard
; Yeh, Patrice
; TITLE OF INVENTION: NOVEL BIOLOGICALLY ACTIVE POLYPEPTIDES,
; PREPARATION THEREOF AND PHARMACEUTICAL COMPOSITION
; CONTAINING SAID POLYPEPTIDES
; NUMBER OF SEQUENCES: 36
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Rhone-Poulenc Rorer Inc.
; STREET: 500 Arcola Road, 3C43
; CITY: Collegeville
; STATE: PA
; COUNTRY: USA
; ZIP: 19426
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: Macintosh
; OPERATING SYSTEM: System 7.1
; SOFTWARE: Word 5.1 (PatentIn)
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/984,186
; FILING DATE: 29-Oct-2001
; CLASSIFICATION: <Unknown>
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US/08/797,689
; FILING DATE: 31-JAN-1997
; APPLICATION NUMBER: US 08/256,927
; FILING DATE: 28-JUL-1994
; APPLICATION NUMBER: FR 92/01064
; FILING DATE: 31-JAN-1992
; APPLICATION NUMBER: PCT/FR93/00085
; FILING DATE: 28-JAN-1993
; ATTORNEY/AGENT INFORMATION:
; NAME: Smith Ph.D., Julie K.
; REGISTRATION NUMBER: P-38 619
; REFERENCE/DOCKET NUMBER: ST92006-US
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (610) 454-3839
; TELEFAX: (610) 454-3808
; INFORMATION FOR SEQ ID NO: 11:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 423 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: cDNA
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 3..419
; SEQUENCE DESCRIPTION: SEQ ID NO: 11:

```

```

US-09-984-186-11
Query Match 72.3%; Score 18.8; DB 4; Length 423;
Best Local Similarity 90.9%; Pred. No. 11;
Matches 20; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

OY 5 CTGGAAGTCTGCCAGTGTCTT 26
Db 230 CTGGAAGTCTGCCAGTGTCTT 251

RESULT 3
US-09-101-272G-72
; Sequence 72, Application US/09101272G
; Patent No. 6509445
; GENERAL INFORMATION:
; APPLICANT: Nissin Food Products Co., Ltd.
; TITLE OF INVENTION: CANCEROUS METASTASIS INHIBITOR
; FILE REFERENCE: Q50979
; CURRENT APPLICATION NUMBER: US/09/101,272G
; CURRENT FILING DATE: 1998-07-08
; PRIOR APPLICATION NUMBER: JP 1059/1996
; PRIOR FILING DATE: 1996-01-08
; NUMBER OF SEQ ID NOS: 107
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 72
; LENGTH: 600
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc feature
; OTHER INFORMATION: ATF domain of uPA
; NAME/KEY: CDS
; LOCATION: (1)..(600)
; OTHER INFORMATION:
; NAME/KEY: mat_peptide
; LOCATION: (61)..(1)
; OTHER INFORMATION:
; US-09-101-272G-72
Query Match 72.3%; Score 18.8; DB 4; Length 600;
Best Local Similarity 90.9%; Pred. No. 12;
Matches 20; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

OY 5 CTGGAAGTCTGCCAGTGTCTT 26
Db 279 CTGGAAGTCTGCCAGTGTCTT 300

RESULT 4
US-09-101-272G-79
; Sequence 79, Application US/09101272G
; Patent No. 6509445
; GENERAL INFORMATION:
; APPLICANT: Nissin Food Products Co., Ltd.
; TITLE OF INVENTION: CANCEROUS METASTASIS INHIBITOR
; FILE REFERENCE: Q50979
; CURRENT APPLICATION NUMBER: US/09/101,272G
; CURRENT FILING DATE: 1998-07-08
; PRIOR APPLICATION NUMBER: JP 1059/1996
; PRIOR FILING DATE: 1996-01-08
; NUMBER OF SEQ ID NOS: 107
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 79
; LENGTH: 624
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: ATFHI chimeric protein
; NAME/KEY: CDS
; LOCATION: (12)..(593)
; OTHER INFORMATION:
; NAME/KEY: mat_peptide

```

```
; NAME/KEY: mat_peptide
; LOCATION: (15)..()
; OTHER INFORMATION:
US-09-101-272G-95

Query Match      72.3%; Score 18.8; DB 4; Length 624;
Best Local Similarity 90.9%; Pred. No. 12;
Matches 20; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 5 CTGGAAGTCTGCCAGTGTCTT 26
Db 233 CTGGAAGTCTGCCAGTGTCTT 254

RESULT 5
US-09-101-272G-95
; Sequence 95, Application US/09101272G
; Patent No. 6509445
; GENERAL INFORMATION:
; APPLICANT: Nissin Food Products Co., Ltd.
; TITLE OF INVENTION: CANCEROUS METASTASIS INHIBITOR
; FILE REFERENCE: Q50979
; CURRENT APPLICATION NUMBER: US/09/101,272G
; CURRENT FILING DATE: 1998-07-08
; PRIOR APPLICATION NUMBER: JP 1059/1996
; PRIOR FILING DATE: 1996-01-08
; NUMBER OF SEQ ID NOS: 107
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 95
; LENGTH: 645
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: ATPHI-CL chimeric protein
; NAME/KEY: CDS
; LOCATION: (12)..(614)
; OTHER INFORMATION:
; NAME/KEY: mat_peptide
; LOCATION: (15)..()
; OTHER INFORMATION:
US-09-101-272G-95

Query Match      72.3%; Score 18.8; DB 4; Length 645;
Best Local Similarity 90.9%; Pred. No. 12;
Matches 20; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 5 CTGGAAGTCTGCCAGTGTCTT 26
Db 233 CTGGAAGTCTGCCAGTGTCTT 254

RESULT 6
US-09-101-272G-97
; Sequence 97, Application US/09101272G
; Patent No. 6509445
; GENERAL INFORMATION:
; APPLICANT: Nissin Food Products Co., Ltd.
; TITLE OF INVENTION: CANCEROUS METASTASIS INHIBITOR
; FILE REFERENCE: Q50979
; CURRENT APPLICATION NUMBER: US/09/101,272G
; CURRENT FILING DATE: 1998-07-08
; PRIOR APPLICATION NUMBER: JP 1059/1996
; PRIOR FILING DATE: 1996-01-08
; NUMBER OF SEQ ID NOS: 107
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 97
; LENGTH: 666
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: ATPHI-ML chimeric protein
; NAME/KEY: CDS
; LOCATION: (12)..(635)
; OTHER INFORMATION:

; NAME/KEY: mat_peptide
; LOCATION: (15)..()
; OTHER INFORMATION:
US-09-101-272G-97

Query Match      72.3%; Score 18.8; DB 4; Length 666;
Best Local Similarity 90.9%; Pred. No. 12;
Matches 20; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 5 CTGGAAGTCTGCCAGTGTCTT 26
Db 233 CTGGAAGTCTGCCAGTGTCTT 254

RESULT 7
US-08-254-922-1
; Sequence 1, Application US/08254922
; Patent No. 5626841
; GENERAL INFORMATION:
; APPLICANT: Victor Gurewich
; TITLE OF INVENTION: USE OF INTRA-PLATELET
; TITLE OF INVENTION: UROKINASE-TYPE PLASMINOGEN
; TITLE OF INVENTION: ACTIVATORS FOR LONG-TERM
; TITLE OF INVENTION: INHIBITION OF THROMBOSIS
; NUMBER OF SEQUENCES: 1
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Fish & Richardson
; STREET: 225 Franklin Street
; CITY: Boston
; STATE: Massachusetts
; COUNTRY: U.S.A.
; ZIP: 02110-2804
; COMPUTER READABLE FORM:
; MEDIUM TYPE: 3.5" Diskette, 1.44 Mb
; COMPUTER: IBM PS/2 Model 50Z or 55SX
; OPERATING SYSTEM: MS-DOS (Version 5.0)
; SOFTWARE: WordPerfect (Version 5.1)
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/254,922
; FILING DATE: June 7, 1994
; CLASSIFICATION: 424
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/014,207 -
; FILING DATE: February 5, 1993
; ATTORNEY/AGENT INFORMATION:
; NAME: J. Peter Fasse
; REGISTRATION NUMBER: 32,983
; REFERENCE/DOCKET NUMBER: 04353/004002
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (617) 542-5070
; TELEFAX: (617) 542-8906
; TELEX: 200154
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1233
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
US-08-254-922-1

Query Match      72.3%; Score 18.8; DB 1; Length 1233;
Best Local Similarity 90.9%; Pred. No. 14;
Matches 20; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 5 CTGGAAGTCTGCCAGTGTCTT 26
Db 219 CTGGAAGTCTGCCAGTGTCTT 240

RESULT 8
US-08-286-748B-1
; Sequence 1, Application US/08286748B
; Patent No. 5759542
```

```

; GENERAL INFORMATION:
; APPLICANT: Victor Gurewicz
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE DELIVERY
; OF DRUGS BY PLATELETS FOR THE TREATMENT OF
; TITLE OF INVENTION: CARDIOVASCULAR AND OTHER DISEASES
; NUMBER OF SEQUENCES: 18
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Fish & Richardson
; STREET: 225 Franklin Street
; CITY: Boston
; STATE: Massachusetts
; COUNTRY: U.S.A.
; ZIP: 02110-2804
; COMPUTER READABLE FORM:
; MEDIUM TYPE: 3.5" Diskette, 1.44 Mb
; OPERATING SYSTEM: MS-DOS (Version 5.0)
; SOFTWARE: WordPerfect (Version 5.1)
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/286,748B
; FILING DATE: August 5, 1994
; CLASSIFICATION: 424
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER:
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: J. Peter Fasse
; REGISTRATION NUMBER: 32,983
; REFERENCE/DOCKET NUMBER: 04547/013001
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (617) 542-5070
; TELEFAX: (617) 542-8906
; TELEX: 200154
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1233
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; US-08-286-748B-1

Query Match 72.3%; Score 18.8; DB 1; Length 1233;
Best Local Similarity 90.9%; Pred. No. 14;
Matches 20; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 5 CTGGAAGTCGCCAGTGTCTT 26
Db 219 CTGGAAGTCGCCAGTGTCTT 240

RESULT 9
US-07-957-039A-7
; Sequence 7, Application US/07957039A
; Patent No. 5389538
; GENERAL INFORMATION:
; APPLICANT: TANABE, TOSHIZUMI
; APPLICANT: MORITA, MASANORI
; APPLICANT: HIROSE, MASAAKI
; APPLICANT: ANASUII, YASUO
; TITLE OF INVENTION: MUTANT HUMAN PROUROKINASE
; NUMBER OF SEQUENCES: 8
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Sughrue, Mion, Zinn, Macpeak & Seas
; STREET: 2100 Pennsylvania Avenue
; CITY: Washington
; STATE: DC
; COUNTRY: USA
; ZIP: 20037
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; OPERATING SYSTEM: IBM PC compatible
; SOFTWARE: PC-DOS/MS-DOS
; PATENTIN RELEASE #1.0, Version #1.25

```

```

; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/07/957,039A
; FILING DATE: 06-OCT-1992
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: JP 289257/1991
; FILING DATE: 07-OCT-1991
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (202)293-7060
; TELEFAX: (202)293-7860
; TELEX: 6491103
; INFORMATION FOR SEQ ID NO: 7:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1236 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: both
; MOLECULE TYPE: DNA (genomic)
; ORIGINAL SOURCE:
; INDIVIDUAL ISOLATE: human
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 1..1233
; US-07-957-039A-7

Query Match 72.3%; Score 18.8; DB 1; Length 1236;
Best Local Similarity 90.9%; Pred. No. 14;
Matches 20; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 5 CTGGAAGTCGCCAGTGTCTT 26
Db 219 CTGGAAGTCGCCAGTGTCTT 240

RESULT 10
US-08-153-799-17
; Sequence 17, Application US/08153799
; Patent No. 5768883
; GENERAL INFORMATION:
; APPLICANT: Ballance, David J
; APPLICANT: Goodey, Andrew R
; TITLE OF INVENTION: Polypeptides
; NUMBER OF SEQUENCES: 23
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: R Hain Swope, BOC Health Care Inc
; STREET: 100 Mountain Avenue
; CITY: Murray Hill
; STATE: New Jersey
; COUNTRY: USA
; ZIP: 07974
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; OPERATING SYSTEM: IBM PC compatible
; SOFTWARE: Patent In Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/153,799
; FILING DATE:
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/847975
; FILING DATE: 06-MAR-1992
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: GB 8909916.2
; FILING DATE: 29-APR-1989
; PRIOR APPLICATION DATA: PCT/GB90/00650
; APPLICATION NUMBER:
; FILING DATE: 26-APR-1990
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/775952
; FILING DATE: 29-OCT-1991
; ATTORNEY/AGENT INFORMATION:
; NAME: Swope, R Hain

```

```
/
/ REGISTRATION NUMBER: 24864
/ REFERENCE/DOCKET NUMBER: 92H832
/ TELECOMMUNICATION INFORMATION:
/ TELEPHONE: (908) 665 2400
/ TELEFAX: (908) 771 6159
/ TELEX: 219484
/ INFORMATION FOR SEQ ID NO: 17:
/ SEQUENCE CHARACTERISTICS:
/ LENGTH: 1236 base pairs
/ TYPE: nucleic acid
/ STRANDEDNESS: double
/ TOPOLOGY: linear
/ MOLECULE TYPE: DNA (genomic)
/ HYPOTHETICAL: NO
/ ANTI-SENSE: NO
/ ORIGINAL SOURCE:
/ ORGANISM: Homo sapiens
/ FEATURE:
/ NAME/KEY: CDS
/ LOCATION: 1..1236
/ OTHER INFORMATION: /function= "human mature
/ OTHER INFORMATION: urokinase-type plasminogen activator (uPA)"
/ FEATURE:
/ NAME/KEY: primer_bind
/ LOCATION: 13..47
/ OTHER INFORMATION: /standard_name= "PCR primer binding
/ OTHER INFORMATION: site"
/ FEATURE:
/ NAME/KEY: primer_bind
/ LOCATION: 376..418
/ OTHER INFORMATION: /standard_name= "PCR primer binding
/ OTHER INFORMATION: site"
/
/ US-08-153-799-17
/
/ Query Match 72.3%; Score 18.8; DB 1; Length 1236;
/ Best Local Similarity 90.9%; Pred. No. 14;
/ Matches 20; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
/
/ QY 5 CTGGAAGTCTGCCAGTGTCTT 26
/ Db 219 CTGGAAGTCTGCCAGTGTCTT 240
/
/ RESULT 11
/ US-09-023-655-927
/ Sequence 927, Application US/09023655
/ Patent No. 6607879
/ GENERAL INFORMATION:
/ APPLICANT: Cocks, Benjamin G.
/ APPLICANT: Susan G. Stuart
/ APPLICANT: Jeffrey J. Seilhamer
/ TITLE OF INVENTION: COMPOSITION FOR THE DETECTION OF BLOOD CELL GENE
/ TITLE OF INVENTION: EXPRESSION
/ NUMBER OF SEQUENCES: 1508
/ CORRESPONDENCE ADDRESS:
/ ADDRESSEE: INCYTE PHARMACEUTICALS, INC.
/ STREET: 3174 PORTER DRIVE
/ CITY: PALO ALTO
/ STATE: CALIFORNIA
/ COUNTRY: USA
/ ZIP: 94304
/ COMPUTER READABLE FORM:
/ MEDIUM TYPE: Floppy disk
/ COMPUTER: IBM PC compatible
/ OPERATING SYSTEM: PC-DOS/MS-DOS
/ SOFTWARE: Word Perfect 6.1 for Windows/MS-DOS 6.2
/ CURRENT APPLICATION DATA:
/ APPLICATION NUMBER: US/09/023,655
/ FILING DATE: HEREWITH
/ CLASSIFICATION:
/ PRIOR APPLICATION DATA:
/ APPLICATION NUMBER:
/ FILING DATE:
/
/ US-09-023-655-927
/ Query Match 72.3%; Score 18.8; DB 4; Length 1236;
/ Best Local Similarity 90.9%; Pred. No. 14;
/ Matches 20; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
/
/ QY 5 CTGGAAGTCTGCCAGTGTCTT 26
/ Db 219 CTGGAAGTCTGCCAGTGTCTT 240
/
/ RESULT 12
/ 5219569-1
/ Patent No. 5219569
/ APPLICANT: BLABER, MICHAEL; HEYNEKER, HERBERT L.; VEHAR,
/ GORDON A.
/ TITLE OF INVENTION: PROTEASE RESISTANT UROKINASE
/ NUMBER OF SEQUENCES: 6
/ CURRENT APPLICATION DATA:
/ APPLICATION NUMBER: US/07/766,858
/ FILING DATE: 16-AUG-1985
/ PRIOR APPLICATION DATA:
/ APPLICATION NUMBER: 725,468
/ FILING DATE: 22-APR-1985
/ SEQ ID NO: 1:
/ LENGTH: 1372
/ 5219569-1
/
/ Query Match 72.3%; Score 18.8; DB 6; Length 1372;
/ Best Local Similarity 90.9%; Pred. No. 14;
/ Matches 20; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
/
/ QY 5 CTGGAAGTCTGCCAGTGTCTT 26
/ Db 355 CTGGAAGTCTGCCAGTGTCTT 376
/
/ RESULT 13
/ US-09-643-597-122
/ Sequence 122, Application US/09643597
/ Patent No. 6426072
/ GENERAL INFORMATION:
/ APPLICANT: Wang, Tongtong
/ APPLICANT: Fan, Liqun
/ APPLICANT: Kalos, Michael D.
/ APPLICANT: Bangur, Chaitanya S.
/ APPLICANT: Hosken, Nancy
/ APPLICANT: Fanger, Gary R.
/ APPLICANT: Li, Samuel X.
/ APPLICANT: Wang, Aijun
/ APPLICANT: Skeiky, Yasir A.W.
/ APPLICANT: Henderson, Robert A.
/ APPLICANT: McNeill, Patricia D.
/ TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY
/ TITLE OF INVENTION: AND DIAGNOSIS OF LUNG CANCER
/ FILE REFERENCE: 210121.455C11
```



; CURRENT APPLICATION NUMBER: US/09/643,597  
; CURRENT FILING DATE: 2000-08-21  
; NUMBER OF SEQ ID NOS: 369  
; SOFTWARE: FastSeq for Windows Version 3.0  
; SEQ ID NO 122  
; LENGTH: 1475  
; TYPE: DNA  
; ORGANISM: Homo sapien  
US-09-643-597-122

Query Match 72.3%; Score 18.8; DB 4; Length 1475;  
Best Local Similarity 90.9%; Pred. No. 14;  
Matches 20; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 5 CTGGAAGTCTGCCAGTGTCTT 26  
Db 359 CTGGAAGTCTGCCAGTGTCTT 380

## RESULT 14

US-09-480-884A-122  
; Sequence 122, Application US/09480884A  
; Patent No. 6482597  
; GENERAL INFORMATION:  
; APPLICANT: Wang, Tongtong  
; APPLICANT: Fan, Liqun  
; APPLICANT: Hosken, Nancy A.  
; APPLICANT: Kalos, Michael D.  
; APPLICANT: Fanger, Gary R.  
; TITLE OF INVENTION: COMPOUNDS AND METHODS FOR THERAPY  
; TITLE OF INVENTION: AND DIAGNOSIS OF LUNG CANCER  
; FILE REFERENCE: 210121.455C6  
; CURRENT APPLICATION NUMBER: US/09/480,884A  
; CURRENT FILING DATE: 2001-08-27  
; NUMBER OF SEQ ID NOS: 330  
; SOFTWARE: FastSeq for Windows Version 3.0  
; SEQ ID NO 122  
; LENGTH: 1475  
; TYPE: DNA  
; ORGANISM: Homo sapien  
US-09-480-884A-122

Query Match 72.3%; Score 18.8; DB 4; Length 1475;  
Best Local Similarity 90.9%; Pred. No. 14;  
Matches 20; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 5 CTGGAAGTCTGCCAGTGTCTT 26  
Db 359 CTGGAAGTCTGCCAGTGTCTT 380

## RESULT 15

US-09-542-615A-122  
; Sequence 122, Application US/09542615A  
; Patent No. 6518256  
; GENERAL INFORMATION:  
; APPLICANT: Wang, Tongtong  
; APPLICANT: Fan, Liqun  
; APPLICANT: Kalos, Michael D.  
; APPLICANT: Bangur, Chaitanya S.  
; APPLICANT: Hosken, Nancy A.  
; APPLICANT: Fanger, Gary R.  
; TITLE OF INVENTION: COMPOUNDS AND METHODS FOR THERAPY  
; TITLE OF INVENTION: AND DIAGNOSIS OF LUNG CANCER  
; FILE REFERENCE: 210121.455C8  
; CURRENT APPLICATION NUMBER: US/09/542,615A  
; CURRENT FILING DATE: 2000-04-14  
; NUMBER OF SEQ ID NOS: 350  
; SOFTWARE: FastSeq for Windows Version 3.0  
; SEQ ID NO 122  
; LENGTH: 1475  
; TYPE: DNA  
; ORGANISM: Homo sapien

US-09-542-615A-122

Query Match 72.3%; Score 18.8; DB 4; Length 1475;  
Best Local Similarity 90.9%; Pred. No. 14;  
Matches 20; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 5 CTGGAAGTCTGCCAGTGTCTT 26  
Db 359 CTGGAAGTCTGCCAGTGTCTT 380

Search completed: February 20, 2004, 01:45:11  
Job time : 6.21902 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: February 19, 2004, 23:23:34 ; Search time 52.4956 Seconds  
(without alignments)  
1456.787 Million cell updates/sec

Title: US-09-939-853A-141  
Perfect score: 26  
Sequence: 1 CCTCTGGAAGTCCAGTGCCTT 26

Scoring table: IDENTITY NUC  
Gapop 10.0 , Gapext 1.0

Searched: 2308684 seqs, 1750822206 residues

Total number of hits satisfying chosen parameters: 4617368

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : Published Applications NA.\*

- 1: /cgn2\_6/ptodata/2/pubna/US07\_PUBCOMB.seq.\*
- 2: /cgn2\_6/ptodata/2/pubna/US06\_NEW\_PUB.seq.\*
- 3: /cgn2\_6/ptodata/2/pubna/US06\_NEW\_PUB.seq.\*
- 4: /cgn2\_6/ptodata/2/pubna/US06\_PUBCOMB.seq.\*
- 5: /cgn2\_6/ptodata/2/pubna/US07\_NEW\_PUB.seq.\*
- 6: /cgn2\_6/ptodata/2/pubna/US07\_PUBCOMB.seq.\*
- 7: /cgn2\_6/ptodata/2/pubna/US08\_NEW\_PUB.seq.\*
- 8: /cgn2\_6/ptodata/2/pubna/US08\_PUBCOMB.seq.\*
- 9: /cgn2\_6/ptodata/2/pubna/US09A\_PUBCOMB.seq.\*
- 10: /cgn2\_6/ptodata/2/pubna/US09B\_PUBCOMB.seq.\*
- 11: /cgn2\_6/ptodata/2/pubna/US09C\_PUBCOMB.seq.\*
- 12: /cgn2\_6/ptodata/2/pubna/US09\_NEW\_PUB.seq.\*
- 13: /cgn2\_6/ptodata/2/pubna/US10A\_PUBCOMB.seq.\*
- 14: /cgn2\_6/ptodata/2/pubna/US10B\_PUBCOMB.seq.\*
- 15: /cgn2\_6/ptodata/2/pubna/US10C\_PUBCOMB.seq.\*
- 16: /cgn2\_6/ptodata/2/pubna/US10\_NEW\_PUB.seq.\*
- 17: /cgn2\_6/ptodata/2/pubna/US60\_NEW\_PUB.seq.\*
- 18: /cgn2\_6/ptodata/2/pubna/US60\_PUBCOMB.seq.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
C 1	26	100.0	444	9	US-09-867-550-951 Sequence 951, App
C 2	26	100.0	763	9	US-09-867-550-953 Sequence 953, App
C 3	26	100.0	864	10	US-09-814-353-21302 Sequence 21302, A
4	20.4	78.5	611	15	US-10-027-632-195852 Sequence 195852, A
5	18.8	72.3	258	15	US-10-233-675A-2 Sequence 2, Appli
6	18.8	72.3	258	15	US-10-233-675A-6 Sequence 6, Appli
7	18.8	72.3	258	15	US-10-233-675A-8 Sequence 8, Appli
8	18.8	72.3	258	15	US-10-233-675A-23 Sequence 23, Appl
9	18.8	72.3	258	15	US-10-233-675A-26 Sequence 26, Appl
10	18.8	72.3	264	9	US-09-880-503-10 Sequence 10, Appl
11	18.8	72.3	288	9	US-09-880-503-18 Sequence 18, Appl
12	18.8	72.3	405	9	US-09-880-503-13 Sequence 13, Appl
13	18.8	72.3	423	9	US-09-984-186-11 Sequence 11, Appl
14	18.8	72.3	423	14	US-10-237-867-11 Sequence 11, Appl
15	18.8	72.3	423	14	US-10-237-708-11 Sequence 11, Appl

16	18.8	72.3	423	14	US-10-237-866-11 Sequence 11, Appl
17	18.8	72.3	423	14	US-10-237-871-11 Sequence 11, Appl
18	18.8	72.3	423	14	US-10-237-624-11 Sequence 11, Appl
19	18.8	72.3	482	9	US-09-880-503-17 Sequence 17, Appl
20	18.8	72.3	482	9	US-09-917-800A-581 Sequence 581, Appl
21	18.8	72.3	1212	9	US-09-880-503-15 Sequence 15, Appl
22	18.8	72.3	1236	9	US-09-880-503-12 Sequence 12, Appl
23	18.8	72.3	1236	15	US-10-407-821-1 Sequence 1, Appli
24	18.8	72.3	1296	13	US-10-076-421-1 Sequence 1, Appli
25	18.8	72.3	1475	9	US-09-735-705-122 Sequence 122, App
26	18.8	72.3	1475	9	US-09-850-116A-122 Sequence 122, App
27	18.8	72.3	1475	9	US-09-897-778-122 Sequence 122, App
28	18.8	72.3	1475	10	US-09-466-396A-122 Sequence 122, App
29	18.8	72.3	1475	14	US-10-117-982-122 Sequence 122, App
30	18.8	72.3	1475	14	US-10-101-510-159 Sequence 159, App
31	18.8	72.3	1475	15	US-10-313-986-122 Sequence 122, App
32	18.8	72.3	1964	14	US-10-131-985-22 Sequence 22, Appli
33	18.8	72.3	1996	15	US-10-401-077-2 Sequence 2, Appli
34	18.8	72.3	2294	9	US-09-735-705-123 Sequence 123, App
35	18.8	72.3	2294	9	US-09-850-116A-123 Sequence 123, App
36	18.8	72.3	2294	9	US-09-897-778-123 Sequence 123, App
37	18.8	72.3	2294	10	US-09-466-396A-123 Sequence 123, App
38	18.8	72.3	2294	14	US-10-117-982-123 Sequence 123, App
39	18.8	72.3	2294	15	US-10-313-986-123 Sequence 123, App
40	18.8	72.3	2304	14	US-10-171-311-183 Sequence 183, App
41	18.8	72.3	2304	14	US-10-301-822-160 Sequence 160, App
42	18.8	72.3	2304	15	US-10-295-027-413 Sequence 413, App
43	18.8	72.3	2304	15	US-10-295-027-1058 Sequence 1058, App
44	18.8	72.3	2304	15	US-10-159-563-121 Sequence 121, App
45	18.8	72.3	2328	14	US-10-247-671-34 Sequence 34, Appli

# ALIGNMENTS

## RESULT 1

US-09-867-550-951/c  
; Sequence 951, Application US/09867550  
; Patent No. US20020082206A1  
; GENERAL INFORMATION:  
; APPLICANT: Leach, Martin D.  
; APPLICANT: Mehraban, Fuad,  
; APPLICANT: Conley, Pamela  
; APPLICANT: Law, Debbie  
; APPLICANT: Topper, James  
; TITLE OF INVENTION: No. US20020082206A1 Polynucleotides from Atherogenic Cells and i  
; FILE REFERENCE: 21402-013 (Cura-313)  
; CURRENT APPLICATION NUMBER: US/09/867,550  
; CURRENT FILING DATE: 2001-09-20  
; PRIOR APPLICATION NUMBER: USSN 60/208,427  
; PRIOR FILING DATE: 2000-05-30  
; NUMBER OF SEQ ID NOS: 2125  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 951  
; LENGTH: 444  
; TYPE: DNA  
; ORGANISM: Homo sapiens  
US-09-867-550-951

Query Match 100.0%; Score 26; DB 9; Length 444;  
Best Local Similarity 100.0%; Pred. No. 0.012;  
Matches 26; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 CCTCTGGAAGTCCAGTGCCTT 26  
|||||  
Db 35 CCTCTGGAAGTCCAGTGCCTT 10

## RESULT 2

US-09-867-550-953/c  
; Sequence 953, Application US/09867550  
; Patent No. US20020082206A1

```

; GENERAL INFORMATION:
; APPLICANT: Leach, Martin D.
; APPLICANT: Mehraban, Fuad.
; APPLICANT: Conley, Pamela
; APPLICANT: Law, Debbie
; APPLICANT: Topper, James
; TITLE OF INVENTION: No. US2002082206A1el Polynucleotides from Atherogenic Cells and
; FILE REFERENCE: 21402-013 (Gura-313)
; CURRENT APPLICATION NUMBER: US/09/867,550
; PRIOR FILING DATE: 2001-09-20
; PRIOR APPLICATION NUMBER: USSN 60/208,427
; PRIOR FILING DATE: 2000-05-30
; NUMBER OF SEQ ID NOS: 2125
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 953
; LENGTH: 763
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-867-550-953

Query Match      100.0%; Score 26; DB 9; Length 763;
Best Local Similarity 100.0%; Pred. No. 0.013;
Matches 26; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 CCTCTGGAAGTCTGCCAGTGCCTT 26
DB 164 CCTCTGGAAGTCTGCCAGTGCCTT 139

RESULT 3
US-09-814-353-21302/c
; Sequence 21302, Application US/09814353
; Publication No. US20030165831A1
; GENERAL INFORMATION:
; APPLICANT: Lee, John
; APPLICANT: Thompson, Pamela
; APPLICANT: Lillie, James
; TITLE OF INVENTION: NOVEL GENES, COMPOSITIONS, KITS, AND METHODS FOR
; TITLE OF INVENTION: IDENTIFICATION, ASSESSMENT, PREVENTION, AND
; FILE REFERENCE: THERAPY OF OVARIAN CANCER
; FILE REFERENCE: MRI-006B
; CURRENT APPLICATION NUMBER: US/09/814,353
; CURRENT FILING DATE: 2001-03-21
; PRIOR APPLICATION NUMBER: US 60/191,031
; PRIOR FILING DATE: 2000-03-21
; PRIOR APPLICATION NUMBER: US 60/207,124
; PRIOR FILING DATE: 2000-05-25
; PRIOR APPLICATION NUMBER: US 60/211,940
; PRIOR FILING DATE: 2000-06-15
; PRIOR APPLICATION NUMBER: US 60/216,820
; PRIOR FILING DATE: 2000-07-07
; PRIOR APPLICATION NUMBER: US 60/220,661
; PRIOR FILING DATE: 2000-07-25
; PRIOR APPLICATION NUMBER: US 60/257,672
; PRIOR FILING DATE: 2000-12-21
; NUMBER OF SEQ ID NOS: 22037
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 21302
; LENGTH: 864
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc.feature
; LOCATION: 1, 2, 3, 32, 862, 863, 864
; OTHER INFORMATION: n = A,T,C or G
US-09-814-353-21302

Query Match      100.0%; Score 26; DB 10; Length 864;
Best Local Similarity 100.0%; Pred. No. 0.013;
Matches 26; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 CCTCTGGAAGTCTGCCAGTGCCTT 26

```

```

DB 328 CCTCTGGAAGTCTGCCAGTGCCTT 303

RESULT 4
US-10-027-632-195852
; Sequence 195852, Application US/10027632
; Publication No. US20030204075A9
; GENERAL INFORMATION:
; APPLICANT: Wang, David G.
; TITLE OF INVENTION: Identification and Mapping of Single Nucleotide
; TITLE OF INVENTION: Polymorphisms in the Human Genome
; FILE REFERENCE: 108827.129
; CURRENT APPLICATION NUMBER: US/10/027,632
; CURRENT FILING DATE: 2002-04-30
; PRIOR APPLICATION NUMBER: US 60/218,006
; PRIOR FILING DATE: 2000-07-12
; PRIOR APPLICATION NUMBER: US 60/198,676
; PRIOR FILING DATE: 2000-04-20
; PRIOR APPLICATION NUMBER: US 60/193,483
; PRIOR FILING DATE: 2000-03-29
; PRIOR APPLICATION NUMBER: US 60/185,218
; PRIOR FILING DATE: 2000-02-24
; PRIOR APPLICATION NUMBER: US 60/167,363
; PRIOR FILING DATE: 1999-11-23
; PRIOR APPLICATION NUMBER: US 60/156,358
; PRIOR FILING DATE: 1999-09-28
; PRIOR APPLICATION NUMBER: US 60/146,002
; PRIOR FILING DATE: 1999-08-09
; NUMBER OF SEQ ID NOS: 325720
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 195852
; LENGTH: 611
; TYPE: DNA
; ORGANISM: Human
US-10-027-632-195852

Query Match      78.5%; Score 20.4; DB 15; Length 611;
Best Local Similarity 95.5%; Pred. No. 5.9;
Matches 21; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 2 CCTCTGGAAGTCTGCCAGTGC 23
DB 484 CCTCTGGAAGTCTGCCAGTGC 505

RESULT 5
US-10-233-675A-2
; Sequence 2, Application US/10233675A
; Publication No. US2003028289A1
; GENERAL INFORMATION:
; APPLICANT: Nesbit, Mark
; APPLICANT: Fong, Timothy
; APPLICANT: Brockstedt, Dirk
; TITLE OF INVENTION: Abrogen Polypeptides, Nucleic Acids Encoding Them and Methods for
; TITLE OF INVENTION: Them To Inhibit Angiogenesis
; FILE REFERENCE: ST01027
; CURRENT APPLICATION NUMBER: US/10/233,675A
; CURRENT FILING DATE: 2002-09-04
; PRIOR APPLICATION NUMBER: 60/316,300
; PRIOR FILING DATE: 2001-09-04
; NUMBER OF SEQ ID NOS: 27
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 2
; LENGTH: 258
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: human derived abrogen
US-10-233-675A-2

Query Match      72.3%; Score 18.8; DB 15; Length 258;
Best Local Similarity 90.9%; Pred. No. 33;

```

Matches 20; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 5 CTGGAAGTCTGCCAGTGTCTT 26  
 Db 78 CTGGAAGTCTGCCAGTGTCTT 99

RESULT 6

US-10-233-675A-6  
 ; Sequence 6, Application US/102333675A  
 ; Publication No. US20030228298A1  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Nesbit, Mark  
 ; APPLICANT: Fong, Timothy  
 ; APPLICANT: Brockstedt, Dirk  
 ; TITLE OF INVENTION: Abrogen Polypeptides, Nucleic Acids Encoding Them and Methods for  
 ; TITLE OF INVENTION: Them To Inhibit Angiogenesis  
 ; FILE REFERENCE: ST01027  
 ; CURRENT APPLICATION NUMBER: US/10/233,675A  
 ; PRIOR FILING DATE: 2002-09-04  
 ; PRIOR APPLICATION NUMBER: 60/316,300  
 ; PRIOR FILING DATE: 2001-09-04  
 ; NUMBER OF SEQ ID NOS: 27  
 ; SOFTWARE: Patent in version 3.1  
 ; SEQ ID NO 6  
 ; LENGTH: 258  
 ; TYPE: DNA  
 ; ORGANISM: Artificial Sequence  
 ; FEATURE:  
 ; OTHER INFORMATION: human derived abrogen  
 ; US-10-233-675A-6

Query Match 72.3%; Score 18.8; DB 15; Length 258;  
 Best Local Similarity 90.9%; Pred. No. 33;  
 Matches 20; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 5 CTGGAAGTCTGCCAGTGTCTT 26  
 Db 78 CTGGAAGTCTGCCAGTGTCTT 99

RESULT 7

US-10-233-675A-8  
 ; Sequence 8, Application US/102333675A  
 ; Publication No. US20030228298A1  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Nesbit, Mark  
 ; APPLICANT: Fong, Timothy  
 ; APPLICANT: Brockstedt, Dirk  
 ; TITLE OF INVENTION: Abrogen Polypeptides, Nucleic Acids Encoding Them and Methods for  
 ; TITLE OF INVENTION: Them To Inhibit Angiogenesis  
 ; FILE REFERENCE: ST01027  
 ; CURRENT APPLICATION NUMBER: US/10/233,675A  
 ; PRIOR FILING DATE: 2002-09-04  
 ; PRIOR APPLICATION NUMBER: 60/316,300  
 ; PRIOR FILING DATE: 2001-09-04  
 ; NUMBER OF SEQ ID NOS: 27  
 ; SOFTWARE: Patent in version 3.1  
 ; SEQ ID NO 8  
 ; LENGTH: 258  
 ; TYPE: DNA  
 ; ORGANISM: Artificial Sequence  
 ; FEATURE:  
 ; OTHER INFORMATION: human derived abrogen  
 ; US-10-233-675A-8

Query Match 72.3%; Score 18.8; DB 15; Length 258;  
 Best Local Similarity 90.9%; Pred. No. 33;  
 Matches 20; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 5 CTGGAAGTCTGCCAGTGTCTT 26  
 Db 78 CTGGAAGTCTGCCAGTGTCTT 99

RESULT 8

US-10-233-675A-23  
 ; Sequence 23, Application US/102333675A  
 ; Publication No. US20030228298A1  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Nesbit, Mark  
 ; APPLICANT: Fong, Timothy  
 ; APPLICANT: Brockstedt, Dirk  
 ; TITLE OF INVENTION: Abrogen Polypeptides, Nucleic Acids Encoding Them and Methods for  
 ; TITLE OF INVENTION: Them To Inhibit Angiogenesis  
 ; FILE REFERENCE: ST01027  
 ; CURRENT APPLICATION NUMBER: US/10/233,675A  
 ; CURRENT FILING DATE: 2002-09-04  
 ; PRIOR APPLICATION NUMBER: 60/316,300  
 ; PRIOR FILING DATE: 2001-09-04  
 ; NUMBER OF SEQ ID NOS: 27  
 ; SOFTWARE: Patent in version 3.1  
 ; SEQ ID NO 23  
 ; LENGTH: 258  
 ; TYPE: DNA  
 ; ORGANISM: Artificial Sequence  
 ; FEATURE:  
 ; OTHER INFORMATION: fragment of human urokinase plasminogen activator  
 ; US-10-233-675A-23

Query Match 72.3%; Score 18.8; DB 15; Length 258;  
 Best Local Similarity 90.9%; Pred. No. 33;  
 Matches 20; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 5 CTGGAAGTCTGCCAGTGTCTT 26  
 Db 78 CTGGAAGTCTGCCAGTGTCTT 99

RESULT 9

US-10-233-675A-26  
 ; Sequence 26, Application US/102333675A  
 ; Publication No. US20030228298A1  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Nesbit, Mark  
 ; APPLICANT: Fong, Timothy  
 ; APPLICANT: Brockstedt, Dirk  
 ; TITLE OF INVENTION: Abrogen Polypeptides, Nucleic Acids Encoding Them and Methods for  
 ; TITLE OF INVENTION: Them To Inhibit Angiogenesis  
 ; FILE REFERENCE: ST01027  
 ; CURRENT APPLICATION NUMBER: US/10/233,675A  
 ; CURRENT FILING DATE: 2002-09-04  
 ; PRIOR APPLICATION NUMBER: 60/316,300  
 ; PRIOR FILING DATE: 2001-09-04  
 ; NUMBER OF SEQ ID NOS: 27  
 ; SOFTWARE: Patent in version 3.1  
 ; SEQ ID NO 26  
 ; LENGTH: 258  
 ; TYPE: DNA  
 ; ORGANISM: Artificial Sequence  
 ; FEATURE:  
 ; OTHER INFORMATION: fragment of human urokinase plasminogen activator  
 ; US-10-233-675A-26

Query Match 72.3%; Score 18.8; DB 15; Length 258;  
 Best Local Similarity 90.9%; Pred. No. 33;  
 Matches 20; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 5 CTGGAAGTCTGCCAGTGTCTT 26  
DB 78 CTGGAAGTCTGCCAGTGTCTT 99

## RESULT 10

US-09-880-503-10  
; Sequence 10, Application US/09880503  
; Patent No. US20020131964A1  
; GENERAL INFORMATION:  
; APPLICANT: CINES, Douglas B  
; APPLICANT: HIGAZI, Abd Al-Roof  
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR MODULATING MUSCLE CELL AND  
; FILE REFERENCE: 9596-331  
; CURRENT APPLICATION NUMBER: US/09/880,503  
; CURRENT FILING DATE: 2001-06-13  
; PRIOR APPLICATION NUMBER: US 60/212,847  
; PRIOR FILING DATE: 2000-06-20  
; NUMBER OF SEQ ID NOS: 18  
; SOFTWARE: PatentIn Ver. 2.1  
; SEQ ID NO 10  
; LENGTH: 264  
; TYPE: DNA  
; ORGANISM: Homo sapiens  
US-09-880-503-10

Query Match 72.3%; Score 18.8; DB 9; Length 264;  
Best Local Similarity 90.9%; Pred. No. 33;  
Matches 20; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 5 CTGGAAGTCTGCCAGTGTCTT 26  
DB 78 CTGGAAGTCTGCCAGTGTCTT 99

## RESULT 11

US-09-880-503-18  
; Sequence 18, Application US/09880503  
; Patent No. US20020131964A1  
; GENERAL INFORMATION:  
; APPLICANT: CINES, Douglas B  
; APPLICANT: HIGAZI, Abd Al-Roof  
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR MODULATING MUSCLE CELL AND  
; FILE REFERENCE: 9596-331  
; CURRENT APPLICATION NUMBER: US/09/880,503  
; CURRENT FILING DATE: 2001-06-13  
; PRIOR APPLICATION NUMBER: US 60/212,847  
; PRIOR FILING DATE: 2000-06-20  
; NUMBER OF SEQ ID NOS: 18  
; SOFTWARE: PatentIn Ver. 2.1  
; SEQ ID NO 18  
; LENGTH: 288  
; TYPE: DNA  
; ORGANISM: Homo sapiens  
US-09-880-503-18

Query Match 72.3%; Score 18.8; DB 9; Length 288;  
Best Local Similarity 90.9%; Pred. No. 33;  
Matches 20; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 5 CTGGAAGTCTGCCAGTGTCTT 26  
DB 78 CTGGAAGTCTGCCAGTGTCTT 99

## RESULT 12

US-09-880-503-13  
; Sequence 13, Application US/09880503  
; Patent No. US20020131964A1  
; GENERAL INFORMATION:  
; APPLICANT: CINES, Douglas B

; APPLICANT: HIGAZI, Abd Al-Roof  
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR MODULATING MUSCLE CELL AND  
; FILE REFERENCE: 9596-331  
; CURRENT APPLICATION NUMBER: US/09/880,503  
; CURRENT FILING DATE: 2001-06-13  
; PRIOR APPLICATION NUMBER: US 60/212,847  
; PRIOR FILING DATE: 2000-06-20  
; NUMBER OF SEQ ID NOS: 18  
; SOFTWARE: PatentIn Ver. 2.1  
; SEQ ID NO 13  
; LENGTH: 405  
; TYPE: DNA  
; ORGANISM: Homo sapiens  
US-09-880-503-13

Query Match 72.3%; Score 18.8; DB 9; Length 405;  
Best Local Similarity 90.9%; Pred. No. 34;  
Matches 20; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 5 CTGGAAGTCTGCCAGTGTCTT 26  
DB 219 CTGGAAGTCTGCCAGTGTCTT 240

## RESULT 13

US-09-984-186-11  
; Sequence 11, Application US/09984186  
; Patent No. US20020151011A1  
; GENERAL INFORMATION:  
; APPLICANT: Fleer, Reinhard  
; Fournier, Alain  
; Guittou, Jean-Dominique  
; Jung, Gerard  
; Yeh, Patrice  
; TITLE OF INVENTION: NOVEL BIOLOGICALLY ACTIVE POLYPEPTIDES,  
; PREPARATION THEREOF AND PHARMACEUTICAL COMPOSITION  
; CONTAINING SAID POLYPEPTIDES

NUMBER OF SEQUENCES: 36  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Rhone-Poulenc Rorer Inc.  
STREET: 500 Arcola Road, 3C43  
CITY: Collegeville  
STATE: PA  
COUNTRY: USA  
ZIP: 19426

COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: Macintosh  
OPERATING SYSTEM: System 7.1  
SOFTWARE: Word 5.1 (Patentin)  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/09/984,186  
FILING DATE: 29-Oct-2001  
CLASSIFICATION: <Unknown>

PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US/08/797,689  
FILING DATE: 31-JAN-1997  
APPLICATION NUMBER: US 08/256,927  
FILING DATE: 28-JUL-1994  
APPLICATION NUMBER: FR 92/01064  
FILING DATE: 31-JAN-1992  
APPLICATION NUMBER: PCT/FR93/00085  
FILING DATE: 28-JAN-1993

ATTORNEY/AGENT INFORMATION:  
NAME: Smith, Ph.D., Julie K.  
REGISTRATION NUMBER: P-38,619  
REFERENCE/DOCKET NUMBER: ST92006-US  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (610) 454-3839  
TELEFAX: (610) 454-3808  
INFORMATION FOR SEQ ID NO: 11:  
SEQUENCE CHARACTERISTICS:

```

;
; LENGTH: 423 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: cdna
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 3..419
; SEQUENCE DESCRIPTION: SEQ ID NO: 11:
US-09-984-186-11

Query Match      72.3%; Score 18.8; DB 9; Length 423;
Best Local Similarity 90.9%; Pred. No. 34;
Matches 20; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 5 CTGGAAGTCTGCCAGTGTCTT 26
Db 230 CTGGAAGTCTGCCAGTGTCTT 251

RESULT 14
US-10-237-667-11
; Sequence 11, Application US/10237667
; Publication No. US20030022308A1
; GENERAL INFORMATION:
; APPLICANT: Fleer, Reinhard
; Fournier, Alain
; Guitton, Jean-Dominique
; Jung, Gerard
; Yeh, Patrice
;
; TITLE OF INVENTION: NOVEL BIOLOGICALLY ACTIVE POLYPEPTIDES,
; PREPARATION THEREOF AND PHARMACEUTICAL COMPOSITION
; CONTAINING SAID POLYPEPTIDES
;
; NUMBER OF SEQUENCES: 36
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Rhone-Poulenc Rorer Inc.
; STREET: 500 Arcola Road, 3C43
; CITY: Collegeville
; STATE: PA
; COUNTRY: USA
; ZIP: 19426
;
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: Macintosh
; OPERATING SYSTEM: System 7.1
; SOFTWARE: Word 5.1 (Patentin)
;
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/10/237,667
; FILING DATE: 10-Sep-2002
; CLASSIFICATION: <Unknown>
;
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US/08/797,689
; FILING DATE: 31-JAN-1997
; APPLICATION NUMBER: US 08/256,927
; FILING DATE: 28-JUL-1994
; APPLICATION NUMBER: FR 92/01064
; FILING DATE: 31-JAN-1992
; APPLICATION NUMBER: PCT/FR93/00085
; FILING DATE: 28-JAN-1993
;
; ATTORNEY/AGENT INFORMATION:
; NAME: Smith Ph.D., Julie K.
; REGISTRATION NUMBER: P-38,619
; REFERENCE/DOCKET NUMBER: ST92006-US
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (610) 454-3839
; TELEFAX: (610) 454-3808
;
; INFORMATION FOR SEQ ID NO: 11:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 423 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: cdna
;
; NAME/KEY: CDS
; LOCATION: 3..419
; SEQUENCE DESCRIPTION: SEQ ID NO: 11:
US-10-237-708-11

;
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 3..419
; SEQUENCE DESCRIPTION: SEQ ID NO: 11:
US-10-237-667-11

Query Match      72.3%; Score 18.8; DB 14; Length 423;
Best Local Similarity 90.9%; Pred. No. 34;
Matches 20; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 5 CTGGAAGTCTGCCAGTGTCTT 26
Db 230 CTGGAAGTCTGCCAGTGTCTT 251

RESULT 15
US-10-237-708-11
; Sequence 11, Application US/10237708
; Publication No. US20030036170A1
; GENERAL INFORMATION:
; APPLICANT: Fleer, Reinhard
; Fournier, Alain
; Guitton, Jean-Dominique
; Jung, Gerard
; Yeh, Patrice
;
; TITLE OF INVENTION: NOVEL BIOLOGICALLY ACTIVE POLYPEPTIDES,
; PREPARATION THEREOF AND PHARMACEUTICAL COMPOSITION
; CONTAINING SAID POLYPEPTIDES
;
; NUMBER OF SEQUENCES: 36
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Rhone-Poulenc Rorer Inc.
; STREET: 500 Arcola Road, 3C43
; CITY: Collegeville
; STATE: PA
; COUNTRY: USA
; ZIP: 19426
;
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: Macintosh
; OPERATING SYSTEM: System 7.1
; SOFTWARE: Word 5.1 (Patentin)
;
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/10/237,708
; FILING DATE: 10-Sep-2002
; CLASSIFICATION: <Unknown>
;
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US/08/797,689
; FILING DATE: 31-JAN-1997
; APPLICATION NUMBER: US 08/256,927
; FILING DATE: 28-JUL-1994
; APPLICATION NUMBER: FR 92/01064
; FILING DATE: 31-JAN-1992
; APPLICATION NUMBER: PCT/FR93/00085
; FILING DATE: 28-JAN-1993
;
; ATTORNEY/AGENT INFORMATION:
; NAME: Smith Ph.D., Julie K.
; REGISTRATION NUMBER: P-38,619
; REFERENCE/DOCKET NUMBER: ST92006-US
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (610) 454-3839
; TELEFAX: (610) 454-3808
;
; INFORMATION FOR SEQ ID NO: 11:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 423 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: cdna
;
; NAME/KEY: CDS
; LOCATION: 3..419
; SEQUENCE DESCRIPTION: SEQ ID NO: 11:
US-10-237-708-11
```

Query Match 72.3%; Score 18.8; DB 14; Length 423;  
Best Local Similarity 90.9%; Pred. No. 34;  
Matches 20; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Cy 5 CTGGAAGTCTGCCAGTGCCTT 26  
Db 230 CTGGAAGTCTGCCAGTGCCTT 251

Search completed: February 20, 2004, 05:59:21  
Job time : 64.4956 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: February 19, 2004, 21:21:18 ; Search time 96.3181 Seconds  
(without alignments)  
9899.970 Million cell updates/sec

Title: US-09-939-853A-142

Perfect score: 22  
Sequence: 1 tgaagagagttctgggtgtrccta 22

Scoring table: IDENTIFY NUC  
Gapop 10.0 , Gapext 1.0

Searched: 3470272 seqs, 2167151695 residues

Total number of hits satisfying chosen parameters: 6940544

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000  
Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database :

GenEmbl.\*

- 1: gb.ba.\*
- 2: gb.htg.\*
- 3: gb.in.\*
- 4: gb.om.\*
- 5: gb.ov.\*
- 6: gb.pat.\*
- 7: gb.ph.\*
- 8: gb.pl.\*
- 9: gb.pr.\*
- 10: gb.ro.\*
- 11: gb.sts.\*
- 12: gb.sy.\*
- 13: gb.un.\*
- 14: gb.vi.\*
- 15: em.ba.\*
- 16: em.fun.\*
- 17: em.hum.\*
- 18: em.in.\*
- 19: em.mu.\*
- 20: em.om.\*
- 21: em.or.\*
- 22: em.ov.\*
- 23: em.pat.\*
- 24: em.pl.\*
- 25: em.pr.\*
- 26: em.ro.\*
- 27: em.sts.\*
- 28: em.un.\*
- 29: em.vi.\*
- 30: em.htg.hum.\*
- 31: em.htg.inv.\*
- 32: em.htg.other.\*
- 33: em.htg.mus.\*
- 34: em.htg.pln.\*
- 35: em.htg.rcd.\*
- 36: em.htg.nam.\*
- 37: em.htg.vrt.\*
- 38: em.sy.\*
- 39: em.htgo.hum.\*
- 40: em.htgo.mus.\*
- 41: em.htgo.other.\*

Pred. No. is the number of results predicted by chance to have a

score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	22	100.0	22	6	AX443201 Sequence
2	22	100.0	1183	6	AX443133 Sequence
3	22	100.0	1183	6	AX443135 Sequence
4	22	100.0	2538	9	BC042041 Homo sapi
5	22	100.0	2567	6	AX452880 Sequence
6	22	100.0	2788	6	AX780857 Sequence
7	22	100.0	65741	9	HS450J78
8	22	100.0	145833	2	AC026539 Homo sapi
9	19.4	88.2	110900	9	AC005079_2
10	19.4	88.2	170128	2	AC005059 Homo sapi
11	19.4	88.2	224064	2	AC005055 Homo sapi
12	18.8	85.5	3521	3	AY058623 Drosophila
13	18.8	85.5	26121	2	AC015346 Drosophila
14	18.8	85.5	177917	9	AC025824 Homo sapi
15	18.8	85.5	180213	3	AC010847 Drosophila
16	18.8	85.5	182623	3	AC011071 Drosophila
17	18.8	85.5	186540	9	AC012478 Homo sapi
18	18.8	85.5	186674	9	AC091544 Homo sapi
19	18.8	85.5	188365	2	AC116758 Mus muscu
20	18.8	85.5	346474	3	AE003512 Drosophila
21	18.4	83.6	76640	10	AL928847 Mouse DNA
22	18.4	83.6	178980	2	AC120340 Mus muscu
23	18.4	83.6	181751	2	AC101938 Mus muscu
24	18.4	83.6	193252	2	AC122504 Mus muscu
25	18.4	83.6	245603	2	AC124616 Mus muscu
26	18	81.8	226598	2	AC099631 Mus muscu
27	18	81.8	271101	2	AC132376 Mus muscu
28	17.8	80.9	1177	3	AK114842 Ciona int
29	17.8	80.9	2064	6	AX834238 Sequence
30	17.8	80.9	2084	9	AK096566 Homo sapi
31	17.8	80.9	45126	9	AL590402 Human DNA
32	17.8	80.9	70328	9	AL513472 Human DNA
33	17.8	80.9	91125	9	AC004899 Homo sapi
34	17.8	80.9	148869	9	AC024093 Homo sapi
35	17.8	80.9	155085	9	AC010412 Homo sapi
36	17.8	80.9	162190	2	AC020592 Homo sapi
37	17.8	80.9	188670	2	AC139753 Mus muscu
38	17.8	80.9	169072	5	AL929396 Zebrafish
39	17.8	80.9	171343	2	AC016676 Homo sapi
40	17.8	80.9	172980	10	AC126671 Mus muscu
41	17.8	80.9	197012	2	AL669823 Mouse DNA
42	17.8	80.9	202681	2	AC111090 Mus muscu
43	17.8	80.9	204131	2	AC128898 Rattus no
44	17.8	80.9	210194	10	AL805896 Mouse DNA
45	17.8	80.9	216589	9	AY129950 Homo sapi

## ALIGNMENTS

RESULT 1  
AX443201  
LOCUS AX443201  
DEFINITION Sequence 142 from Patent WO0216599.  
ACCESSION AX443201  
VERSION AX443201.1 GI:21690596  
KEYWORDS  
SOURCE synthetic construct  
ORGANISM synthetic construct  
artificial sequences.  
REFERENCE 1  
AUTHORS Burgess,C.E., Conley,P.B., Grosse,W.M., Hart,M., Kekuda,R., Shimkets,R.A., Spytek,K.A., Szekeres,E.S., Tomlinson,J.E., Topper,J.N. and Yang,R.B.  
TITLE Proteins and nucleic acids encoding same

22 bp DNA linear PAT 02-JUL-2002



JOURNAL Patent: WO 0216599-A 142 28-FEB-2002;  
Curagen Corporation (US); COR THERAPEUTICS, INC. (US)  
FEATURES Location/Qualifiers  
source 1..22

/organism="synthetic construct"  
/mol\_type="unassigned DNA"  
/db\_xref="taxon:32630"  
/note="oligonucleotide primer"

## ORIGIN

Query Match 100.0%; Score 22; DB 6; Length 22;  
Best Local Similarity 100.0%; Pred. No. 0.27;  
Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 TGAGAGAGTTCTGGGTGTCCTA 22  
|||||  
Db 1 TGAGAGAGTTCTGGGTGTCCTA 22

## RESULT 2

AX443133 1183 bp DNA linear PAT 02-JUL-2002  
LOCUS Sequence 74 from Patent WO0216599.  
DEFINITION AX443133  
ACCESSION AX443133  
VERSION AX443133.1 GI:21690555  
KEYWORDS  
SOURCE Homo sapiens (human)  
ORGANISM Homo sapiens

REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
AUTHORS Burgess,C.E., Conley,P.B., Grosse,W.M., Hart,M., Kekuda,R.,  
Shinkets,R.A., Spytek,K.A., Szekeres,E.S., Tomlinson,J.E.,  
Topper,J.N. and Yang,R.B.  
TITLE Proteins and nucleic acids encoding same  
JOURNAL Patent: WO 0216599-A 74 28-FEB-2002;  
Curagen Corporation (US); COR THERAPEUTICS, INC. (US)

## FEATURES

source 1..1183  
/organism="Homo sapiens"  
/mol\_type="unassigned DNA"  
/db\_xref="taxon:9606"

## ORIGIN

Query Match 100.0%; Score 22; DB 6; Length 1183;  
Best Local Similarity 100.0%; Pred. No. 0.36;  
Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 TGAGAGAGTTCTGGGTGTCCTA 22  
|||||  
Db 224 TGAGAGAGTTCTGGGTGTCCTA 245

## RESULT 3

AX443135/c 1183 bp DNA linear PAT 02-JUL-2002  
LOCUS Sequence 76 from Patent WO0216599.  
DEFINITION AX443135  
ACCESSION AX443135  
VERSION AX443135.1 GI:21690556  
KEYWORDS  
SOURCE Homo sapiens (human)  
ORGANISM Homo sapiens

## REFERENCE

1 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
AUTHORS Burgess,C.E., Conley,P.B., Grosse,W.M., Hart,M., Kekuda,R.,  
Shinkets,R.A., Spytek,K.A., Szekeres,E.S., Tomlinson,J.E.,  
Topper,J.N. and Yang,R.B.  
TITLE Proteins and nucleic acids encoding same  
JOURNAL Patent: WO 0216599-A 76 28-FEB-2002;  
Curagen Corporation (US); COR THERAPEUTICS, INC. (US)

## FEATURES

source 1..1183  
Location/Qualifiers

/organism="Homo sapiens"  
/mol\_type="unassigned DNA"  
/db\_xref="taxon:9606"

Query Match 100.0%; Score 22; DB 6; Length 1183;  
Best Local Similarity 100.0%; Pred. No. 0.36;  
Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 TGAGAGAGTTCTGGGTGTCCTA 22  
|||||  
Db 960 TGAGAGAGTTCTGGGTGTCCTA 939

## RESULT 4

BC042041 2538 bp mRNA linear PRI 07-OCT-2003  
LOCUS Homo sapiens Src-like-adaptor 2, transcript variant 1, mRNA (cDNA  
clone MGC:49845 IMAGE:4429896), complete cds.  
DEFINITION BC042041  
ACCESSION BC042041  
VERSION BC042041.1 GI:27469842  
KEYWORDS MSC  
SOURCE Homo sapiens (human)  
ORGANISM Homo sapiens

## REFERENCE

1 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
AUTHORS Strausberg,R.L., Feingold,E.A., Grouse,L.H., Derge,J.G.,  
Klausner,R.D., Collins,F.S., Wagner,L., Shenmen,C.M., Schuler,G.D.,  
Altschul,S.F., Zeeberg,B., Buetow,K.H., Schaefer,C.F., Bhat,N.K.,  
Hopkins,R.F., Jordan,H., Moore,T., Max,S.I., Wang,J., Hsieh,F.,  
Dhatchenkov,L., Marusina,K., Farmer,A.A., Rubin,G.M., Hong,L.,  
Sapleton,M., Soares,M.B., Bonaldo,M.F., Casavant,T.L.,  
Schectz,T.E., Brownstein,M.J., Usdin,T.B., Toshiyuki,S.,  
Carninci,P., Frange,C., Raha,S.S., Loquellano,N.A., Peters,G.J.,  
Abramson,K.D., Mullaly,S.J., Bosak,S.A., McEwan,P.J.,  
McKernan,K.J., Malek,J.A., Gunaratne,P.H., Richards,S.,  
Worley,K.C., Hale,S., Garcia,A.M., Gay,L.J., Hulyk,S.W.,  
Villalón,D.K., Muzny,D.M., Sodergren,E.J., Lu,X., Gibbs,R.A.,  
Fahey,J., Helton,E., Kettman,M., Madan,A., Rodriguez,S.,  
Sánchez,A., Whiting,M., Madan,A., Young,A.C., Shevchenko,Y.,  
Bouffard,G.G., Blakesley,R.W., Touchman,J.W., Green,E.D.,  
Dickson,M.C., Rodriguez,A.C., Grimwood,J., Schmutz,J., Myers,R.M.,  
Bitterfield,Y.S., Krzywinski,M.I., Skalska,U., Smailus,D.E.,  
Schnerch,A., Schein,J.E., Jones,S.J. and Marra,M.A.

Generation and initial analysis of more than 15,000 full-length  
human and mouse cDNA sequences  
Proc. Natl. Acad. Sci. U.S.A. 99 (26), 16899-16903 (2002)  
22388257  
12477932  
2 (bases 1 to 2538)  
Strausberg,R.  
Direct Submission  
Submitted (23-DEC-2002) National Institutes of Health, Mammalian  
Gene Collection (MGC), Cancer Genomics Office, National Cancer  
Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590,  
USA  
NIH-MGC Project URL: <http://mgc.nci.nih.gov>  
Contact: MGC help desk  
Email: [cgapsb-email.nih.gov](mailto:cgapsb-email.nih.gov)  
Tissue Procurement: DCTD/DTF  
cDNA Library Preparation: Life Technologies, Inc.  
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)  
DNA Sequencing by: Sequencing Group at the Stanford Human Genome  
Center, Stanford University School of Medicine, Stanford, CA 94305  
Web site: <http://www-sngc.stanford.edu>  
Contact: (Dickson, Mark) [mdcpax11.stanford.edu](mailto:mdcpax11.stanford.edu)  
Dickson, M., Schmutz, J., Grimwood, J., Rodriguez, A., and Myers,  
R. M.

## TITLE

JOURNAL  
MEDLINE  
PUBMED  
REFERENCE  
AUTHORS  
TITLE  
JOURNAL

## REMARK

## COMMENT

Clone distribution: MGC clone distribution information can be found  
through the I.M.A.G.E. Consortium/LLNL at: <http://image.llnl.gov>  
Series: IRAX Plate: 88 Row: a Column: 20

This clone was selected for full length sequencing because it passed the following selection criteria: matched mRNA gi: 28416422.

```

FEATURES             source
  source
    1..2538
      /organism="Homo sapiens"
      /mol_type="mRNA"
      /db_xref="taxon:9606"
      /clone="MGC:49845 IMAGE:4429896"
      /tissue types="Prostate, adenocarcinoma."
      /clone_lib="NIH_MGC_91"
      /lab_host="DH10B"
      /notes="Vector: pCMV-SPORT6"
  gene
    1..2538
      /gene="SLA2"
      /notes="synonyms: FLJ21992, SLAP-2, C20orf156, MGC49845,
      SLAP2"
      /db_xref="LocusID:84174"
      /db_xref="MIM:606577"
  CDS
    363..1148
      /codon_start=1
      /product="Src-like adaptor 2, isoform a"
      /protein_id="AAH42041.1"
      /db_xref="GI:27469843"
      /db_xref="LocusID:84174"
      /translation="MGSLSRSRKSLPSPSLSSVQGGPVTVNEARSKATAVALGSFP
      AGFAPLSRLGEPITIVBEGDMTVLSESGREYNIPSVHVAKVSHGWLYEGLSRE
      KAEELLLPQNGPAGFLIBESQTRGSYLSVRLSRPASWDRIRHYRHCLDNGWLYI
      SPLTPPSQALVHYHSEIADICLLKPCVLRAGPLPGKDILPVTVQRTPLNWK
      ELDSLLFSFEATGERSLLSEGLSLSFYISLNDENVSLDDA"
      /misc_feature
        468..632
          /notes="SH3; Region: SH3 domain. SH3 (src homology 3)
          domains are often indicative of a protein involved in
          signal transduction related to cytoskeletal organization.
          First described in the Src cytoplasmic tyrosine kinase.
          The structure is a partly opened beta barrel"
          /db_xref="CDD:pfam00018"
      /misc_feature
        642..890
          /notes="SH2; Region: SH2 domain"
          /db_xref="CDD:pfam00017"
  ORIGIN
    Query Match      100.0%; Score 22; DB 9; Length 2538;
    Best Local Similarity 100.0%; Pred. No. 0.38;
    Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

    QY      1 TGAGAGAGTTCTGGGTGTCCTA 22
      |||||
    Db      189 TGAGAGAGTTCTGGGTGTCCTA 210

  RESULT 5
  AX452880
  LOCUS
  DEFINITION
  Sequence 1 from Patent WO0242457.
  ACCESSION
  AX452880
  VERSION
  AX452880.1 GI:21712520
  SOURCE
  Homo sapiens (human)
  ORGANISM
  Homo sapiens
  Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
  Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
  1
  Chang, H., Yang, W.P., Wu, Y., Whitney, G.S., Perez-Villar, J.J. and
  Kanher, S.B.
  Cloning and expression of human slap-2: a novel sh2/sh3
  domain-containing human slap homologue having immune cell-specific
  expression
  Patent: WO 0242457-A 1 30-MAY-2002;
  Bristol-Myers Squibb Co. (US)
  Location/Qualifiers
    1..2567
      /organism="Homo sapiens"
      /mol_type="unassigned DNA"

  This clone was selected for full length sequencing because it passed the following selection criteria: matched mRNA gi: 28416422.

  ORIGIN
    Query Match      100.0%; Score 22; DB 6; Length 2567;
    Best Local Similarity 100.0%; Pred. No. 0.38;
    Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

    QY      1 TGAGAGAGTTCTGGGTGTCCTA 22
      |||||
    Db      189 TGAGAGAGTTCTGGGTGTCCTA 210

  RESULT 6
  AX780857
  LOCUS
  DEFINITION
  Sequence 3014 from Patent WO03039443.
  ACCESSION
  AX780857
  VERSION
  AX780857.1 GI:32697851
  SOURCE
  Homo sapiens (human)
  ORGANISM
  Homo sapiens
  Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
  Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
  1
  Haerlach, T., Schoch, C., Kern, W., Kohlmann, A., Schnittger, S.,
  Dugas, M., Eils, R., Brors, B. and Meigenthaler, S.
  Novel genetic markers for leukemias
  Patent: WO 03039443-A 3014 15-MAY-2003;
  Deutsches Krebsforschungszentrum (DE);
  Ludwig-Maximilian-Universitaet Muenchen (DE);
  PD Dr. Dr. (DE); Schoch, Claudia (DE); Kern, Wolfgang (DE)
  Location/Qualifiers
    1..2788
      /organism="Homo sapiens"
      /mol_type="unassigned DNA"
      /db_xref="taxon:9606"

  ORIGIN
    Query Match      100.0%; Score 22; DB 6; Length 2788;
    Best Local Similarity 100.0%; Pred. No. 0.38;
    Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

    QY      1 TGAGAGAGTTCTGGGTGTCCTA 22
      |||||
    Db      213 TGAGAGAGTTCTGGGTGTCCTA 234

  RESULT 7
  HS460J8
  LOCUS
  DEFINITION
  Human DNA sequence from clone Rp3-460J8 on chromosome
  20q11.21-11.23 Contains the 3' end of the gene for a novel protein
  similar to N-myc downstream regulated (NDRG1) the 5' end of a gene
  encoding a novel protein tyrosine kinase, ESTs, STSs and GSSs,
  complete sequence.
  AL031662
  VERSION
  AL031662.26 GI:9716901
  KEYWORDS
  HIG; NDRG1; SH2 domain.
  SOURCE
  Homo sapiens (human)
  ORGANISM
  Homo sapiens
  Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
  Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
  1
  (bases 1 to 66741)
  Skuce, C
  Direct Submission
  Submitted (23-JUL-2001) Sanger Centre, Hinxton, Cambridgeshire,
  CB10 1SA, UK. E-mail enquiries: humquery@sanger.ac.uk
  requests: clonerequest@sanger.ac.uk
  On Aug 7, 2000 this sequence version replaced gi:6425549.
  During sequence assembly data is compared from overlapping clones.
  Where differences are found these are annotated as variations
  together with a note of the overlapping clone name. Note that the
  variation annotation may not be found in the sequence submission

```

corresponding to the overlapping clone, as we submit sequences with only a small overlap as described above.

The following abbreviations are used to associate primary accession numbers given in the feature table with their source databases: Em., EMBL; Sw., SWISSPROT; Tr., TREMBL; Wp., WORMPEP; Information on the WORMPEP database can be found at [http://www.sanger.ac.uk/projects/C\\_elegans/wormpep](http://www.sanger.ac.uk/projects/C_elegans/wormpep) This sequence was generated from part of bacterial clone contigs of human chromosome 20, constructed by the Sanger Centre Chromosome 20 Mapping Group. Further information can be found at <http://www.sanger.ac.uk/HGP/Chr20>

IMPORTANT: This sequence is not the entire insert of clone RP3-460J8 It may be shorter because we sequence overlapping sections only once, except for a 100 base overlap.

The true right end of clone RP3-460J8 is at 66741 in this sequence. The left end of clone RP3-469A13 is at 41767 in this sequence. The true right end of clone RP5-977B1 is at 100 in this sequence. This sequence was finished as follows unless otherwise noted: all regions were either double-stranded or sequenced with an alternate chemistry or covered by high quality data (i.e., phred quality >= 30); an attempt was made to resolve all sequencing problems, such as compressions and repeats; all regions were covered by at least one plasmid subclone or more than one M13 subclone; and the assembly was confirmed by restriction digest. RP3-460J8 is from the library RPCI-3 constructed by the group of Pieter de Jong. For further details see <http://www.chori.org/bacpac/home.htm>

VECTOR: pCYPAC2.

## FEATURES

```

source
    1. 66741
       Location/Qualifiers
         /organism="Homo sapiens"
         /mol_type="genomic DNA"
         /db_xref="taxon:9606"
         /chromosome="20"
         /map="q11.21-11.23"
         /clone="RP3-460J8"
         /clone_lib="RPCI-3"
         complement(50..544)
         /note="match: GSS: Em:B45150"
         complement(240..7128)
         /gene="dJ977B1.1"
         complement(join(<240..339,6995..7128))
         /gene="dJ977B1.1"
         /product="dJ460J8.2 (novel protein tyrosine kinase with Src homology 2 (SH2) domain)"
         /note="match: cDNAs: Em:AK025645"
         match: ESTs: Em:BG178487
         /evidence=not_experimental
         complement(join(<240..339,6995..7085))
         /gene="dJ977B1.1"
         /note="Continues in Em:AL050318 as dJ977B1.2"
         /codon_start=1
         /evidence=not_experimental
         /product="dJ460J8.2 (novel protein tyrosine kinase with Src homology 2 (SH2) domain)"
         /protein_id="CAC44645.1"
         /db_xref="GI:15020830"
         /db_xref="GOA:Q9HEQ3"
         /db_xref="SWISS-PROT:Q9HEQ3"
         /translation="VGSUPSRKSLSPSLSSVQGGPVTMEARSKATAVALGSPF
         AGPAAELSLRGLGEPITIVS"
         536..565
         /note="15 copies 2 mer ag 93% conserved"
         2375..2402
         /note="14 copies 2 mer ta 100% conserved"
         2783..3155
         /note="match: GSS: Em:AQ807191"
         complement(6993..7482)
         /note="match: GSS: Em:AQ556467"
         complement(6985..7492)
         /note="match: GSS: Em:AQ556478"
         7486..7925
         /note="match: STS: Em:HS427J15"

```

```

repeat_region
repeat_region
repeat_region
repeat_region
repeat_region
repeat_region
polyA_site
misc_feature
polyA_site
polyA_site
gene
mRNA
polyA_signal
misc_feature
misc_feature
CDS
misc_feature
misc_feature
repeat_region
repeat_region
repeat_region

```

```

10615..10640
/note="13 copies 2 mer tt 92% conserved"
13593..13961
/note="MER21B repeat: matches 422..785 of consensus"
14248..14297
/note="MER21B repeat: matches 374..422 of consensus"
14594..14614
/note="MER21B repeat: matches 355..374 of consensus"
14746..15078
/note="MER21B repeat: matches 1..355 of consensus"
complement(17516)
17519..17686
/note="match: GSS: Em:AZ067993"
complement(17521)
complement(17522)
complement(17523..54534)
/gene="dJ469A13.3"
complement(join(<17523..19451,20551..20589,22110..22161,
22247..22282,26087..26134,30796..30847,30986..31051,
32039..32142,37091..37147,46566..46652,48268..48328,
50163..50225,53242..53362,54429..54534))
/gene="dJ469A13.3"
/products="dJ460J8.1 (continued from dJ469A13.3 in
Em:AL132768)"
/note="match: cDNAs: Em:AB033922 Em:U52073 Em:D87953
Em:X92845 Em:AF004162 Em:U60593 Em:AF045564 Em:AB033921
Em:AF159092 Em:AF147402 Em:M59814
match: ESTs: Em:AA718726 Em:AA039000 Em:AW003952 Em:W89263
Em:AA325826 Em:AI230982 Em:AA162360 Em:AA445016
Em:AV002395 Em:FB5147 Em:AI786615 Em:AA113437 Em:AI004026
Em:T89705 Em:AI786673 Em:AI181197 Em:AA764653 Em:AA316771
Em:AU035165 Em:AV002368"
/evidence=not_experimental
complement(17542..17547)
/gene="dJ469A13.3"
17555..17811
/note="match: STS: Em:G19945"
join(19202..19451,22110..22282,26087..26127,30796..30847)
/note="match: STS: Em:G23762"
complement(join(19270..19451,20551..20589,22110..22161,
22247..22282,26087..26134,30796..30847,30986..31051,
32039..32142,37091..37147,46566..46652,48268..48328,
50163..50225,53242..53362,54429..54534))
/gene="dJ469A13.3"
/note="novel protein (FLJ13556) similar to N-myc
downstream regulated (NDRG1)
match: proteins: Sw:Q62433 Sw:Q92219 Tr:Q92219 Sw:P97862"
/codon_start=1
/evidence=not_experimental
/product="dJ460J8.1 (continued from dJ469A13.3 in
Em:AL132768)"
/protein_id="CAB65625.1"
/db_xref="GI:6687781"
/db_xref="GOA:Q9UGV2"
/db_xref="SWISS-PROT:Q9UGV2"
/translation="EHDITHTGVHTVIRGLPKGNRPVILTYHDIGLNHKSQNAFF
NFEDMQEITQHFVCHVDAPGOEQGAPSPPTGQYQVPTMDELAEM:PPVLTLSLSKII
G:GVGAGVYILSRFALNHPVAVGLINVDPCAKGIDWAKSLSGITNVVDILA
HIFGREGIQLNDLIQTTRMIAQDINQNLQFLNDRKSLDIERPILQGNOKS
YFLKQSTLIVVDNSPAVEAVVCCNSRLNPINTLLRKWADCGGLPQVQPGKLTAFK
YFGMGXIPYVQLSHLSTESVPSASMTLARSRTHSTSSSLGSGESPFSRSVTSNQS
DGTQSCSPVDLDRHQTMEVSC"
complement(24246..24780)
/gene="dJ469A13.3"
/note="match: GSS: Em:AQ592789"
complement(27980..28556)
/gene="dJ469A13.3"
/note="match: GSS: Em:AQ308867"
34313..34420
/note="HY1 repeat: matches 1..109 of consensus"
34505..34560
/note="28 copies 2 mer ta 78% conserved"
34564..34611

```

```

/note="24 copies 2 mer ta 79% conserved"
42553. .43658
/note="MER52C repeat: matches 1. .1278 of consensus"
47989. .48470
/note="match: GSS: Em:AQ672417"
49391. .49480
/note="45 copies 2 mer ta 76% conserved"
52922. .53120
/note="match: STS: Em:G04621"
complement(53426. .54032)
/genes="du469a13.3"
/note="match: GSS: Em:AQ314824
match: STS: Em:G55853"
54054. .54477
/note="match: GSS: Em:AQ310681"
57509. .57838
/note="match: STS: Em:G07504"
complement(57696. .58013)
/note="match: STS: Em:G07632 Em:G07634"
58856. 59001
/note="73 copies 2 mer at 83% conserved"
complement(58930. .59281)
/note="match: GSS: Em:AQ067563"
59062. .59236
/note="TTGGRR2 repeat: matches 2541. .2714 of consensus"
59738. .60495
/note="TTGGRR2 repeat: matches 1780. .2541 of consensus"
60719. .61831
/note="MER11C repeat: matches 1. .1071 of consensus"
62547. .63174

Query Match 100.0%; Score 22; DB 9; Length 66741;
Best Local Similarity 100.0%; Pred NO. 0.47; 0; Indels 0; Gaps 0;
Matches 22; Conservative 0; Mismatches 0;

QY 1 TGAGAGGTTCTGGGTGCTTA 22
Db 11717 TGAGAGGTTCTGGGTGCTTA 11696

RESULT 8
AC026539 145833 bp DNA linear HTG 27-APR-2000
LOCUS Homo sapiens chromosome 20 clone RP11-712N14 map 20, WORKING DRAFT
DEFINITION SEQUENCE, 37 unordered pieces.
ACCESSION AC026539
VERSION AC026539.2 GI:7656813
KEYWORDS HTG; HTGS PHASE1; HTGS_DRAFT.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 145833)
Birren,B., Linton,L., Nusbaum,C., Lander,E., Abraham,H., Allen,N.,
Anderson,S., Balgwin,J., Barna,N., Bastien,V., Beda,F.,
Boguslavskiy,L., Boukhgalter,B., Brown,A., Burkett,G.,
Campopiano,A., Castle,A., Choepel,Y., Colangelo,M., Collins,S.,
Collamore,A., Cooke,P., Dearellano,K., Dewar,K., Diaz,J.S.,
Dodge,S., Domino,M., Doyle,M., Ferreira,P., Fitzhugh,W., Gage,D.,
Galagan,J., Gardyna,S., Girde,S., Goyette,M., Graham,L.,
Grand-pierre,N., Grant,G., Hagos,B., Jones,C., Kann,L., Karatas,A.,
Howland,J.C., Iliev,I., Johnson,R., Landers,T., Lehoczy,J.,
Klein,J., LaRocque,K., Lamazares,R., Landers,T., Lehoczy,J.,
Levine,R., Lieu,C., Liu,G., Locke,K., Macdonald,P., Marquis,N.,
McCarthy,M., McEwan,P., McGurk,A., McKernan,K., Mcpheeters,R.,
Melarim,J., Meneus,L., Mihova,T., Miranda,C., Mlenga,V., Morrow,J.,
Murphy,T., Naylor,J., Norman,C.H., O'Connor,T., O'Donnell,P.,
O'Neill,D., Olivari,T.M., Oliver,J., Peterson,K., Pierre,N.,
Pisani,C., Pollara,V., Raymond,C., Riley,R., Rogov,P., Rothman,D.,

```

ROY, A., Santos, R., Schauer, S., Severy, P., Spencer, B.,  
 Stange-Thomann, N., Stojanovic, N., Subramanian, A., Talamas, J.,  
 Tesfaye, S., Theodore, J., Tirrell, A., Travers, M., Trigilio, J.,  
 Vassiliev, H., Viel, R., Vo, A., Wilson, B., Wu, X., Wyman, D., Ye, W. J.,  
 Young, G., Zainoun, J., Zimmer, A. and Zody, M.  
 Direct Submission  
 Submitted (22-MAR-2000) Whitehead Institute/MIT Center for Genome  
 Research, 320 Charles Street, Cambridge, MA 02141, USA  
 On Apr 27, 2000 this sequence version replaced gi:7283243.  
 All repeats were identified using RepeatMasker:  
 Smit, A.F.A. & Green, P. (1996-1997)  
 http://ftp.genome.washington.edu/RM/RepeatMasker.html

----- Genome Center

Center: Whitehead Institute/ MIT Center for Genome Research  
 Center code: WIBR

Web site: http://www-seq.wi.mit.edu

Contact: sequence\_submissions@genome.wi.mit.edu

----- Project Information

Center project name: L7115

Center clone name: 712 N.14

----- Summary Statistics

Sequencing vector: M13; M77815; 100% of reads

Assembly: Dye-terminator Big Dye; 100% of reads

Assembly program: Phrap; version 0.960731

Consensus quality: 135577 bases at least Q40

Consensus quality: 135703 bases at least Q30

Consensus quality: 139593 bases at least Q20

Insert size: 182000; agarose-fp

Insert size: 142233; sum-of-contigs

Quality coverage: 2.6 in Q20 bases; agarose-fp

Quality coverage: 3.3 in Q20 bases; sum-of-contigs

-----  
 \* NOTE: This is a 'working draft' sequence. It currently  
 \* consists of 37 contigs. The true order of the pieces  
 \* is not known and their order in this sequence record is  
 \* arbitrary. Gaps between the contigs are represented as  
 \* runs of N, but the exact sizes of the gaps are unknown.  
 \* This record will be updated with the finished sequence  
 \* as soon as it is available and the accession number will  
 \* be preserved.

1 1214: contig of 1214 bp in length  
 1215 1314: gap of 100 bp  
 1315 2673: contig of 1359 bp in length  
 2674 2773: gap of 100 bp  
 2774 4520: contig of 1747 bp in length  
 4521 4620: gap of 100 bp  
 4621 5962: contig of 1341 bp in length  
 5962 6062: gap of 100 bp  
 6062 7719: contig of 1658 bp in length  
 7720 7820: gap of 100 bp  
 7820 9899: contig of 1980 bp in length  
 9899 9900: gap of 100 bp  
 9900 11434: contig of 1535 bp in length  
 11435 11534: gap of 100 bp  
 11535 14382: contig of 2848 bp in length  
 14383 14482: gap of 100 bp  
 14483 17148: contig of 2666 bp in length  
 17149 17248: gap of 100 bp  
 17249 19239: contig of 1991 bp in length  
 19239 19339: gap of 100 bp  
 19340 21102: contig of 1763 bp in length  
 21103 21202: gap of 100 bp  
 21203 23371: contig of 2169 bp in length  
 23372 23471: gap of 100 bp  
 23472 25782: contig of 2311 bp in length  
 25783 25882: gap of 100 bp  
 25883 28924: contig of 2942 bp in length  
 28925 31619: contig of 2635 bp in length  
 31620 31719: gap of 100 bp  
 31720 34679: contig of 2960 bp in length  
 34680 34779: gap of 100 bp  
 34780 37527: contig of 2748 bp in length

```

* 37528 37627: gap of 100 bp
* 37628 40146: contig of 2519 bp in length
* 40147 40246: gap of 100 bp
* 4247 43743: contig of 3497 bp in length
* 43744 43843: gap of 100 bp
* 47133 47233: contig of 3290 bp in length
* 47134 51023: gap of 100 bp
* 47234 51023: contig of 3790 bp in length
* 51024 51123: gap of 100 bp
* 51124 54935: contig of 3812 bp in length
* 54936 55035: gap of 100 bp
* 55036 55533: contig of 4518 bp in length
* 55534 59653: gap of 100 bp
* 59654 64428: contig of 4775 bp in length
* 64429 64529: gap of 100 bp
* 64529 69211: contig of 4683 bp in length
* 69212 72011: gap of 100 bp
* 72012 72011: contig of 3590 bp in length
* 72012 73001: gap of 100 bp
* 73001 76724: contig of 3723 bp in length
* 76725 76824: gap of 100 bp
* 76825 81179: contig of 4355 bp in length
* 81180 81279: gap of 100 bp
* 81280 87009: contig of 5730 bp in length
* 87010 87109: gap of 100 bp
* 87110 90855: contig of 3746 bp in length
* 90856 90955: gap of 100 bp
* 90956 96520: contig of 5565 bp in length
* 96521 96620: gap of 100 bp
* 96621 102321: contig of 5701 bp in length
* 102322 102421: gap of 100 bp
* 102422 108293: contig of 5872 bp in length
* 108294 108393: gap of 100 bp
* 108394 116689: contig of 8296 bp in length
* 116690 116789: gap of 100 bp
* 116790 125264: contig of 8475 bp in length
* 125265 125364: gap of 100 bp
* 125365 136356: contig of 10990 bp in length
* 136356 138355: gap of 100 bp
* 138355 145833: contig of 9379 bp in length.

```

# FEATURES source

```

1..145833
/organism="Homo sapiens"
/mol_type="genomic DNA"
/db_xref="taxon:9606"
/chromosome="20"
/map="20"
/clone="RP11-712N14"
/db_lib="RP11-11 Human Male BAC"
1..1214
/note="assembly_fragment"
1315..2673
/note="assembly_fragment"
2774..4520
/note="assembly_fragment"
4621..5961
/note="assembly_fragment"
6062..7719
/note="assembly_fragment"
7820..9799
/note="assembly_fragment"
9900..11434
/note="assembly_fragment"
11535..14382
/note="assembly_fragment"
14483..17148
/note="assembly_fragment"
17249..19239
/note="assembly_fragment"
19340..21102
/note="assembly_fragment"
21203..23371
/note="assembly_fragment"

```

```

misc_feature 23472..25782
/note="assembly_fragment"
misc_feature 25883..28824
/note="assembly_fragment"
misc_feature 28925..31619
/note="assembly_fragment"
misc_feature 31720..34679
/note="assembly_fragment"
misc_feature 34780..37527
/note="assembly_fragment"
misc_feature 37628..40146
/note="assembly_fragment"
misc_feature 40247..43743
/note="assembly_fragment"
misc_feature 43844..47133
/note="assembly_fragment"
misc_feature 47234..51023
/note="assembly_fragment"
misc_feature 51124..54935
/note="assembly_fragment"
clone_end:T7
vector_side:right"
misc_feature 55036..59553

```

Query Match 100.0%; Score 22; DB 2; Length 145833;  
 Best Local Similarity 100.0%; Pred. No. 0.5;  
 Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 TGAGAGAGTTCTGGGTGTCCTA 22  
 |||||

Db 99968 TGAGAGAGTTCTGGGTGTCCTA 99989  
 |||||

## RESULT 9

AC005079.2

## WPCOMMENT

Sequence split into 5 fragments LOCUS AC005079 Accession AC005079

Fragment Name	Begin	End
AC005079.0	1	110000
AC005079.1	100001	210000
AC005079.2	200001	310000
AC005079.3	300001	410000
AC005079.4	400001	415268

Continuation (3 of 5) of AC005079 from base 200001 (AC005079 Homo sapiens BAC clone CTA-:

Query Match 88.2%; Score 19.4; DB 9; Length 110000;  
 Best Local Similarity 95.2%; Pred. No. 15;  
 Matches 20; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 TGAGAGAGTTCTGGGTGTCCT 21  
 |||||

Db 89559 TGAGAGAGTTCTGGGTGTCCT 89579  
 |||||

## RESULT 10

AC005059/c

## LOCUS

AC005059 170128 bp DNA linear HTG 13-MAR-1999  
 Homo sapiens clone RG074A24, \*\*\* SEQUENCING IN PROGRESS \*\*\*, 25  
 unordered pieces.

ACCESSION AC005059.1 GI:3212932

VERSION HTG; HTGS\_PHASE1.

KEYWORDS Homo sapiens (human)

## SOURCE

Homo sapiens  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

1 (bases 1 to 170128)

Waterston,R.H.

TITLE The sequence of Homo sapiens clone

JOURNAL Unpublished

REFERENCE 2 (bases 1 to 170128)

Waterston,R.H.

AUTHORS

TITLE Direct Submission

JOURNAL

Submitted (12-JUN-1998) Genome Sequencing Center, Washington University School of Medicine, 4444 Forest Park Parkway, St. Louis, MO 63108, USA

COMMENT

\* NOTE: This is a 'working draft' sequence. It currently consists of 25 contigs. The true order of the pieces is not known and their order in this sequence record is arbitrary. Gaps between the contigs are represented as runs of N, but the exact sizes of the gaps are unknown. This record will be updated with the finished sequence as soon as it is available and the accession number will be preserved.

1 1602: contig of 1662 bp in length  
1663 1679: gap of unknown length  
1680 3248: contig of 1569 bp in length  
3249 3265: gap of unknown length  
3266 4902: contig of 1637 bp in length  
4903 4919: gap of unknown length  
4920 6551: contig of 1632 bp in length  
6552 6568: gap of unknown length  
6569 8090: contig of 1522 bp in length  
8091 8107: gap of unknown length  
8108 9832: contig of 1725 bp in length  
9833 9849: gap of unknown length  
9850 11534: contig of 1665 bp in length  
11535 11531: gap of unknown length  
11532 13880: contig of 2349 bp in length  
13881 13897: gap of unknown length  
13898 16342: contig of 2445 bp in length  
16343 16359: gap of unknown length  
16360 20698: contig of 4329 bp in length  
20699 20705: gap of unknown length  
20706 24570: contig of 3865 bp in length  
24571 24587: gap of unknown length  
24588 28578: contig of 3991 bp in length  
28579 28595: gap of unknown length  
28596 35314: contig of 6719 bp in length  
35315 35321: gap of unknown length  
35322 39588: gap of unknown length  
39589 46401: contig of 6843 bp in length  
46402 46418: gap of unknown length  
46419 53312: contig of 6894 bp in length  
53320 53329: gap of unknown length  
53330 60281: contig of 6935 bp in length  
60282 66853: contig of 6572 bp in length  
66854 66870: gap of unknown length  
66871 75039: contig of 8169 bp in length  
75040 84782: contig of 9726 bp in length  
84783 84799: gap of unknown length  
84800 97374: contig of 12575 bp in length  
97375 97391: gap of unknown length  
97392 109799: contig of 12408 bp in length  
109800 109816: gap of unknown length  
109817 124892: contig of 15076 bp in length  
124893 124909: gap of unknown length  
124910 143882: contig of 18973 bp in length  
143883 143899: gap of unknown length  
143900 170128: contig of 26229 bp in length.

FEATURES

source  
1. .170128  
/organism="Homo sapiens"  
/mol\_type="genomic DNA"  
/db\_xref="taxon:9606"  
/clone="RG074A24"

ORIGIN

Query Match 88.2%; Score 19.4; DB 2; Length 170128;  
Best Local Similarity 95.2%; Pred. No. 16;  
Matches 20; Conservative 0; Mismatches 1; Indels 0; Gaps 0;  
QY 1 TGAGAGATTCTGGGTGCT 21

Db 48353 TGAGAGATTCTGGGTGCT 48333

RESULT 11

AC005055/c

LOCUS

DEFINITION

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

REFERENCE

AUTHORS

TITLE

JOURNAL

AUTHORS

TITLE

JOURNAL

AUTHORS

TITLE

JOURNAL

AUTHORS

TITLE

JOURNAL

AUTHORS

TITLE

JOURNAL

AUTHORS

TITLE

JOURNAL

AUTHORS

TITLE

JOURNAL

AUTHORS

TITLE

JOURNAL

AUTHORS

TITLE

JOURNAL

AUTHORS

TITLE

JOURNAL

AUTHORS

TITLE

JOURNAL

AUTHORS

TITLE

JOURNAL

AUTHORS

TITLE

JOURNAL

AUTHORS

TITLE

JOURNAL

AUTHORS

TITLE

JOURNAL

AUTHORS

TITLE

JOURNAL

AUTHORS

TITLE

JOURNAL

AUTHORS

TITLE

JOURNAL

AUTHORS

TITLE

JOURNAL

AUTHORS

TITLE

JOURNAL

AUTHORS

AC005055 224064 bp DNA linear HTG 13-MAR-1999  
Homo sapiens clone RG05CN15, \*\*\* SEQUENCING IN PROGRESS \*\*\*, 26  
unordered pieces.

AC005055.1 GI:3212936

HTG: HTGS PHASE1

Homo sapiens (human)

Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

1 (bases 1 to 224064)

Waterston,R.H.

The sequence of Homo sapiens clone

Unpublished

2 (bases 1 to 224064)

Waterston,R.H.

Direct Submission

Submitted (12-JUN-1998) Genome Sequencing Center, Washington

University School of Medicine, 4444 Forest Park Parkway, St. Louis,

MO 63108, USA

COMMENT

\* NOTE: This is a 'working draft' sequence. It currently consists of 26 contigs. The true order of the pieces is not known and their order in this sequence record is arbitrary. Gaps between the contigs are represented as runs of N, but the exact sizes of the gaps are unknown. This record will be updated with the finished sequence as soon as it is available and the accession number will be preserved.

1 1567: contig of 1567 bp in length  
1568 1584: gap of unknown length  
1585 3757: contig of 2173 bp in length  
3758 3774: gap of unknown length  
3775 5645: contig of 1871 bp in length  
5646 5662: gap of unknown length  
5663 8218: contig of 2556 bp in length  
8219 8235: gap of unknown length  
8236 10568: contig of 2333 bp in length  
10569 10585: gap of unknown length  
10586 12673: contig of 2088 bp in length  
12674 12690: gap of unknown length  
12691 14458: contig of 1768 bp in length  
14459 14475: gap of unknown length  
14476 17399: contig of 2924 bp in length  
17400 17416: gap of unknown length  
17417 20845: contig of 3429 bp in length  
20846 20862: gap of unknown length  
20863 24474: contig of 3612 bp in length  
24475 24491: gap of unknown length  
24492 30780: contig of 6289 bp in length  
30781 30797: gap of unknown length  
30798 36366: contig of 5569 bp in length  
36367 36383: gap of unknown length  
36384 41721: contig of 5338 bp in length  
41722 41738: gap of unknown length  
41739 49140: contig of 7402 bp in length  
49141 49157: gap of unknown length  
49158 54878: contig of 5721 bp in length  
54879 54895: gap of unknown length  
54896 60864: contig of 5969 bp in length  
60865 60881: gap of unknown length  
60882 72111: contig of 11230 bp in length  
72112 72128: gap of unknown length  
72129 81775: contig of 9647 bp in length  
81776 81792: gap of unknown length  
81793 92380: contig of 10588 bp in length  
92381 92397: gap of unknown length  
92398 106304: contig of 13907 bp in length

\* 106305 106321: gap of unknown length  
 \* 106322 119808: contig of 13287 bp in length  
 \* 119809 119825: gap of unknown length  
 \* 119826 136119: contig of 16494 bp in length  
 \* 136120 136136: gap of unknown length  
 \* 136137 157328: contig of 21192 bp in length  
 \* 157329 157345: gap of unknown length  
 \* 157346 176995: contig of 19650 bp in length  
 \* 176996 177012: gap of unknown length  
 \* 177013 198965: contig of 21953 bp in length  
 \* 198966 198982: gap of unknown length  
 \* 198983 224064: contig of 25082 bp in length.

## FEATURES

source

1. .224064  
 /organism="Homo sapiens"  
 /mol\_type="genomic DNA"  
 /db\_xref="taxon:9606"  
 /clone="RG050N15"

## ORIGIN

Query Match 88.2%; Score 19.4; DB 2; Length 224064;  
 Best Local Similarity 95.2%; Pred. No. 16;  
 Matches 20; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 TGAGAGAGTTCTGGGTGCT 21

Db 31626 TGAGAGAGTTCTGGGTGCT 31606

## RESULT 12

AY058623

LOCUS Drosophila melanogaster LD31543 full length cDNA. INV 17-OCT-2001

DEFINITION

ACCESSION AY058623  
 VERSION AY058623.1 GI:16198106  
 KEYWORDS FLI CDNA.  
 SOURCE Drosophila melanogaster (fruit fly)  
 ORGANISM Drosophila melanogaster

REFERENCE

AUTHORS Stapleton, M., Brokstein, P., Hong, L., Agbayani, A., Carlson, J.,  
 Champe, M., Chavez, C., Dorsett, V., Farfan, D., Frise, E., George, R.,  
 Gonzalez, M., Guarin, H., Li, P., Liao, G., Miranda, A., Mungall, C. J.,  
 Nunoo, J., Pacleb, J., Paragas, V., Park, S., Phouanavong, S., Wan, K.,  
 Yu, C., Lewis, S. E., Rubin, G. M. and Celisner, S.  
 Direct Submission  
 Submitted (08-OCT-2001) Berkeley Drosophila Genome Project,  
 Lawrence Berkeley National Laboratory, One Cyclotron Road,  
 Berkeley, CA 94720, USA

TITLE

JOURNAL Sequence submitted by:  
 Lawrence Berkeley National Laboratory  
 Berkeley, CA 94720

## COMMENT

This clone was sequenced as part of a high-throughput process to sequence clones from Drosophila Gene Collection 1 (Rubin et al., Science 2000). The sequence has been subjected to integrity checks for sequence accuracy, presence of a polyA tail and contiguity within 100 kb in the genome. Thus we believe the sequence to reflect accurately this particular cDNA clone. However, there are artifacts associated with the generation of cDNA clones that may have not been detected in our initial analyses such as internal priming, priming from contaminating genomic DNA, retained introns due to reverse transcription of unspliced precursor RNAs, and reverse transcriptase errors that result in single base changes. For further information about this sequence, including its location and relationship to other sequences, please visit our web site (<http://fruitfly.berkeley.edu>) or send email to [cdna@fruitfly.berkeley.edu](mailto:cdna@fruitfly.berkeley.edu).

## FEATURES

source

1. .3521  
 /organism="Drosophila melanogaster"

/mol\_type="mRNA"  
 /strain="y; cn bw sp"  
 /db\_xref="taxon:7227"  
 /map="18C8-18C8"  
 1. .3521  
 /gene="mRps14"  
 /notes="alignment with genomic scaffold AE003512. gene does not completely overlap longest ORF"  
 /db\_xref="FLYBASE:FBgn0044030"  
 56. .3475  
 /gene="mRps14"  
 /notes="Longest ORF"  
 /codon\_start=1  
 /product="LD31543p"  
 /protein\_id="AAL13852.1"  
 /db\_xref="GI:16198107"  
 /db\_xref="FLYBASE:FBgn0044030"  
 /translation="MSRIKKKADHASSRRRTQINIEFSLDHLKAEVLGLQAKSRQ  
 VGRSLVENADFWKFGVGYEAMWSTGQVLPRLAEALSMPQVHRSHKLALRLDA  
 DELLOARDQMOSQFRHLLVLDPRGTEKFORIRKLRQTNRLPIARFKDLREAD  
 TSVRVVAGDTGCKSTQVQVLYDFGYSIACTOPRLACVSLCKRVAHELDDYGS  
 RVAFQIRPFRSRKLTNLPITTEGLLRQLAVALNDQYDALILOEIHERNLFDGFL  
 GVTKCLLRBPOLKILMSATINVELFHGFGEGEARLVQVGRUFPPIKRLVLPPL  
 ELKAGATSKRSQRNRIDPAPFQVLSLDQQYPTSERGCVLIFVGVNEISVVEAV  
 HEYATQTHWLVLPLHSGGAIDQSKVFDYAPEGMRKCLVNTINAIISLTVGFRVW  
 DSGVKEMNFDTCKGRLKEFWKSKSADQKRGAGRTGPGVCFLYTAEQNAFEA  
 YPTPEIYRVFLDMLLQVSMGLPDVRAFPFTEAPETERIEQTILAKQHCALSVBEK  
 ITPGRSLANLPVLSIGKMLLGSVFVEQOLLTLAAMLVQNPLTNEAHTDQCVR  
 ERESLESDHGLFTLVLRNVVQLKWRDGTQWCRRLGIEBOREYEVTKLQOQFOR  
 ILBSCGMVSDSDSLTSAERATHEGLRQKAMKRRORFQEPYORKLLKQSGRVA  
 EDEBEEREAQDDMRVDRLPDPQLALLERSRLDRSHSVVLLKILGSGFYQL  
 AISDEFNYCKGGQFFHRLKPFVQLCHPNSQAFHFKLLKTESLLTPKPFYTPAL  
 PLSKRHQLICYSLETHAKPYLNCIRLPAATLLLFSAIDNAGITQIACDGMGLD  
 DLPMPGSMELLERAIERLRWSRLYDKLDDLKSKQSPSHSDRSALMODLVLY  
 MALDVAYAIRLLIPADIKRLYTHQAPSAKLAKENPFAVDPMTNEBKGLNVSEH  
 VVYSLAEQOQWTTAMDATIRABEWQCTRCDFLEFPDVLQVLVHRSCKRRKQAQESRT  
 AKTTSTESSTEVASSSSSSSGGYCYNCKRELRLTLTIILRHKKQCRNNK"

## ORIGIN

Query Match 85.5%; Score 18.8; DB 3; Length 3521;  
 Best Local Similarity 90.9%; Pred. No. 27;  
 Matches 20; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 TGAGAGAGTTCTGGGTGCT 22

Db 1461 TGAAGAGTTCTGGGTGCT 1482

## RESULT 13

AC015346/c

LOCUS

AC015346 26121 bp DNA linear HTG 16-NOV-1999

DEFINITION

AC015346 Drosophila melanogaster, \*\*\* SEQUENCING IN PROGRESS \*\*\*  
 AC015346  
 AC015346.1 GI:6435989

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

Drosophila melanogaster (fruit fly)

Drosophila melanogaster

Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;

Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;

Ephydroidea; Drosophilidae; Drosophila.

REFERENCE

AUTHORS

TITLE

JOURNAL

COMMENT

This sequence was identified as CD1:10210169 by the submitter.

For further information on this sequence e-mail to [fly@celera.com](mailto:fly@celera.com).

\* NOTE: This is a 'working draft' sequence.

\* This sequence will be replaced

\* by the finished sequence as soon as it is available and

\* the accession number will be preserved.

Location/Qualifiers  
 1. .26121

/organism="Drosophila melanogaster"  
/mol\_type="genomic DNA"  
/db\_xref="taxon:7227"

## ORIGIN

Query Match 85.5%; Score 18.8; DB 2; Length 26121;  
Best Local Similarity 90.9%; Pred. No. 31;  
Matches 20; Conservative 0; Mismatches 2; Indels 0; Gaps 0;  
QY 1 TGAGAGAGTCTGGGTGTCCTA 22  
|||||  
Db 5183 TGAAGAGTCTGGGTGTCCTA 5162

## RESULT 14

AC025824/c  
LOCUS AC025824 177917 bp DNA linear PRI 19-MAR-2002  
DEFINITION Homo sapiens chromosome 15 clone RP11-430M3, complete sequence.  
ACCESSION AC025824  
VERSION AC025824.37 GI:19549230  
KEYWORDS HTG.  
SOURCE Homo sapiens (human)

## ORGANISM

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

## REFERENCE

1 (bases 1 to 177917)

## AUTHORS

Abola, A.P., Bruno, D., Conn, L., Della Rosa, M., Faulkner, D.,  
Pederspiel, N., Glukhov, S., Hansen, N., Herman, Z.S., Hyman, R.,  
Mao, J., Komp, C., Kottler, S., Lan, B., Marathe, R., Miranda, M.,  
Morehouse, A.J., Nguyen, M., Oefner, P., Palm, C.J., Ramirez, D.,  
Southwick, A.M., Webb, C., Wilhelmy, J., Yu, S. and Davis, R.W.

## Unpublished

## JOURNAL

## REFERENCE

## AUTHORS

1 (bases 1 to 177917)  
Bruno, D., Conn, L., Della Rosa, M., Faulkner, D., Pederspiel, N.,  
Glukhov, S., Hansen, N., Hyman, R., Mao, J., Marathe, R.,  
Morehouse, A.J., Oefner, P., Palm, C.J., Ramirez, D.,  
Wilhelmy, J., Yu, S. and Davis, R.W.

## Direct Submission

## TITLE

## JOURNAL

## REFERENCE

## AUTHORS

3 (bases 1 to 177917)  
Abola, A.P., Bruno, D., Conn, L., Della Rosa, M., Faulkner, D.,  
Pederspiel, N., Glukhov, S., Hansen, N., Herman, Z.S., Hyman, R.,  
Komp, C., Kottler, S., Lan, B., Mao, J., Marathe, R., Miranda, M.,  
Morehouse, A.J., Nguyen, M., Oefner, P., Palm, C.J., Ramirez, D.,  
Southwick, A.M., Webb, C., Wilhelmy, J., Yu, S. and Davis, R.W.

## Direct Submission

## TITLE

## JOURNAL

## COMMENT

Submitted (16-MAR-2000) DNA Sequencing and Technology Center,  
Stanford University, 855 California Avenue, Palo Alto, CA 94304,  
USA  
3 (bases 1 to 177917)  
Abola, A.P., Bruno, D., Conn, L., Della Rosa, M., Faulkner, D.,  
Pederspiel, N., Glukhov, S., Hansen, N., Herman, Z.S., Hyman, R.,  
Komp, C., Kottler, S., Lan, B., Mao, J., Marathe, R., Miranda, M.,  
Morehouse, A.J., Nguyen, M., Oefner, P., Palm, C.J., Ramirez, D.,  
Southwick, A.M., Webb, C., Wilhelmy, J., Yu, S. and Davis, R.W.

## Direct Submission

## TITLE

## JOURNAL

## COMMENT

Submitted (19-MAR-2002) Genome Technology Center, Stanford  
University, 855 California Avenue, Palo Alto, CA 94304, USA  
On Mar 19, 2002 this sequence version replaced gi:15252013.  
----- Genome Center  
Center: Stanford DNA Sequencing and Technology Development  
Center  
Center code: SDS/DC  
Web site: <http://sequence-www.stanford.edu/group/human/>  
Contact: [hum-info@sequence.stanford.edu](mailto:hum-info@sequence.stanford.edu)  
----- Project Information  
Center project name: 857  
Center clone name: RP11-430M3

## FEATURES

## source

1..177917  
/organism="Homo sapiens"  
/mol\_type="genomic DNA"

/db\_xref="taxon:9606"  
/chromosome="15"  
/clone="RP11-430M3"  
/clone\_lib="RPC1 human BAC library 11"  
128..131  
/note="Target site of Tn10 insertion in BAC. Tn10  
sequence and target site duplication, CATT, have been  
removed from completed sequence."  
23502..23850  
/note="Single-stranded, single chemistry with consensus  
quality below 30."

## ORIGIN

Query Match 85.5%; Score 18.8; DB 9; Length 177917;  
Best Local Similarity 90.9%; Pred. No. 35;  
Matches 20; Conservative 0; Mismatches 2; Indels 0; Gaps 0;  
QY 1 TGAGAGAGTCTGGGTGTCCTA 22  
|||||  
Db 151626 TCAGAGAGTCTGGGTGTCCTA 151605

## RESULT 15

AC010847/c  
LOCUS AC010847 180213 bp DNA linear INV 17-MAR-2001  
DEFINITION Drosophila melanogaster, chromosome X, region 18D-18D, BAC clone  
BACR10M08, complete sequence.  
ACCESSION AC010847  
VERSION AC010847.11 GI:13374650  
KEYWORDS HTG.

## SOURCE

## ORGANISM

## Drosophila melanogaster (fruit fly)

## Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;

## Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;

## Ephydroidea; Drosophilidae; Drosophila.

## 1 (bases 1 to 180213)

## REFERENCE

## AUTHORS

Celniker, S.E., Adams, M.D., Krommiller, B., Tyler, D., Wan, K.H.,  
Holt, R.A., Evans, C.A., Gocayne, J.D., Ananides, P.G., Brandon, R.C.,  
Rogers, Y., An, H., Baldwin, D., Banzon, J., Beeson, K.Y., Busam, D.A.,  
Carlson, J.W., Center, A., Champe, M., Davenport, L.B., Dietz, S.M.,  
Dodson, K., Dorsett, V., Douc, L.E., Doyle, C., Dresnek, D., Farfan, D.,  
Ferrera, S., Frise, E., Galle, R.F., Garg, N.S., George, R.A.,  
Gonzalez, M., Houck, J., Hoskins, R.A., Hostin, D., Howland, T.J.,  
Ibegwam, C., Jalali, M., Kruse, D., Li, P., Mattel, B., Moshrefi, A.,  
McIntosh, T.C., Moy, M., Murphy, B., Nelson, C., Nelson, K.A., Nunoo, J.,  
Paclet, J., Paragas, V., Park, S., Patel, S., Pfeiffer, B.,  
Phouanavong, S., Pittman, G.S., Puri, V., Richards, S., Scheesler, F.,  
Stapleton, M., Strong, R., Svirskaas, R., Tector, C., Williams, S.M.,  
Zaveri, J.S., Smith, H.O., Rubin, G.M. and Venter, J.C.

## Sequencing of Drosophila chromosome X, region 18D-18D

## Unpublished

## 2 (bases 1 to 180213)

## REFERENCE

## AUTHORS

## Celniker, S.E., Agbayani, A., Arcaina, T.T., Baxter, E., Blazej, R.G.,

## Butenhofer, C., Champe, M., Chavez, C., Chew, M., Ciesiolka, L.,

## Doyle, C.M., Farfan, D.E., Galle, R., George, R.A., Harris, N.L.,

## Hoskins, R.A., Houston, K.A., Hummasti, S.R., Karra, K., Kearney, L.,

## Kim, E., Lee, B., Lewis, S., Li, P., Lonotani, M.A., Mazda, P.,

## Moshrefi, A.R., Moshrefi, M., Nixon, K., Paclet, J.M., Park, S.,

## Pfeiffer, B., Poon, L., Sequeira, A., Sethi, H., Snir, E.,

## Svirskaas, R.R., Wan, K.H., Weinburg, T., Zhang, R., Zieran, L.L. and

## Rubin, G.M.

## Direct Submission

## TITLE

## JOURNAL

## COMMENT

## Submitted (24-SEP-1999) Drosophila Genome Center, Lawrence Berkeley

## Laboratory, MS 64-121, Berkeley, CA 94720, USA

## On Mar 17, 2001 this sequence version replaced gi:6563418.

## Sequence submitted by:

## Berkeley Drosophila Genome Project

## Lawrence Berkeley National Laboratory, MS 64-121

## Berkeley, CA 94720

## This sequence was assembled using end sequences from a whole genome

## shotgun and from subclones of this BAC and its neighboring clones.

## For further information about this sequence, including its location

## and relationship to other sequences, please visit our sequence





GenCore version 5.1.6  
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: February 19, 2004, 21:15:48 ; Search time 20.3821 Seconds  
(without alignments)  
4585.415 Million cell updates/sec

Title: US-09-939-853A-142

Perfect score: 22  
Sequence: 1 ttagagagattctgggtgctcta 22

Scoring table: IDENTITY NUC  
Gapop 10.0 , Gapext 1.0

Searched: 3373863 seqs, 2124093041 residues

Total number of hits satisfying chosen parameters: 6747725

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : N Geneseq 29Jan04: \*  
1: Geneseqn1980s: \*  
2: Geneseqn1990s: \*  
3: Geneseqn2000s: \*  
4: Geneseqn2001as: \*  
5: Geneseqn2001bs: \*  
6: Geneseqn2002s: \*  
7: Geneseqn2003as: \*  
8: Geneseqn2003bs: \*  
9: Geneseqn2003cs: \*  
10: Geneseqn2004s: \*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	22	100.0	22	6	ABK61507 Human NOV
2	22	100.0	1183	6	ABK61465 Human CDN
3	22	100.0	2567	6	AAD43380 Human SFC
4	20.4	92.7	763	6	ABQ98870 Human ORF
5	18.8	85.5	3319	4	ABL28809
6	18.8	85.5	5782	4	ABL28808 Drosophila
7	17.4	79.1	2442	7	ABX70675 Human CDN
8	17.4	79.1	2789	6	AAD30568 Human pro
9	17.4	79.1	3076	4	AAS57383 Human ske
10	17.4	79.1	3210	5	AAS90933
11	17.4	79.1	3327	8	ACH03882
12	17.4	79.1	5149	2	AAT32455
13	17.2	78.2	355	4	AAK67913 Human imm
14	17.2	78.2	356	4	AAK67912 Human imm
15	17.2	78.2	359	4	AAK56372 Human imm
16	17.2	78.2	1163	7	ACA49310
17	17.2	78.2	1182	7	ACA51118
18	17.2	78.2	1182	7	ACA51577
19	17.2	78.2	1185	7	ACA35390
20	17.2	78.2	1457	5	AAK93029
21	17.2	78.2	13327	4	ABL02914 Drosophila
22	16.8	76.4	340	7	ADA72333 Rice Gene
23	16.8	76.4	1382	4	AAK73700 Human imm

C	24	16.8	76.4	2394	9	ADB62889	Human cDN
C	25	16.8	76.4	5313	4	ABL06558	Drosophila
	26	16.8	76.4	28772	6	ABK83555	Human cDN
	27	16.4	74.5	701	4	AAF23001	Human pro
	28	16.4	74.5	1191	6	ABL90353	Human pol
	29	16.4	74.5	1273	2	AAI64550	Activated
	30	16.4	74.5	1357	2	AAH68577	Human pro
	31	16.4	74.5	1373	2	AAI64549	Human act
	32	16.4	74.5	1621	2	AAZ20855	Polynucle
	33	16.4	74.5	1621	2	AAZ20855	Human cDN
	34	16.4	74.5	1621	6	ABA90961	Human pol
	35	16.4	74.5	2275	4	ABL19829	Drosophila
	36	16.4	74.5	2857	4	ABL13023	Drosophila
	37	16.4	74.5	2872	4	AAI61360	Human pol
	38	16.4	74.5	3128	6	ABZ35472	Human gen
	39	16.4	74.5	3128	10	ADE77208	Human cDN
	40	16.4	74.5	3986	5	ABV28780	Human pro
	41	16.4	74.5	3986	5	ABV22948	Human pro
	42	16.4	74.5	4665	6	ABK83947	Human cDN
	43	16.4	74.5	5912	4	ABL19828	Drosophila
	44	16.4	74.5	7420	4	ABL13022	Drosophila
	45	16.4	74.5	27684	8	ADA03062	Human KCN

## ALIGNMENTS

RESULT 1	ABK61507	ABK61507 standard; DNA; 22 BP.
XX	AC	ABK61507;
XX	DT	18-JUN-2002 (first entry)
XX	DE	Human NOV13 RT-PCR primer #2.
XX	DE	Human; ss; NOVX; gene therapy; cardiomyopathy; atherosclerosis; cell signal processing disorder; metabolic pathway modulation disorder; diabetes; cancer; adenocarcinoma; lymphoma; prostate cancer; primer; uterus cancer; immune response; graft-versus-host disease; Exon linking; acquired immunodeficiency syndrome; AIDS; asthma; Crohn's disease; hypertension; congenital heart defects; multiple sclerosis; inflammation; Albright hereditary osteodystrophy; reverse transcriptase PCR.
XX	OS	Homo sapiens.
XX	PN	WO200216599-A2.
XX	PD	28-FEB-2002.
XX	PF	27-AUG-2001; 2001WO-US026510.
XX	PR	25-AUG-2000; 2000US-0228191P.
XX	PR	08-FEB-2001; 2001US-0267300P.
XX	PR	20-FEB-2001; 2001US-0269961P.
XX	PR	20-MAR-2001; 2001US-0277337P.
XX	PA	(CURA-) CURAGEN CORP.
XX	PA	(CORT-) COR THERAPEUTICS INC.
XX	PI	Burgess CE, Conley PB, Grosse WM, Hart M, Kekuda R, Shimkets RA, Spytek KA, Szekeres ES, Tomlinson JE, Topper JN, Yang R, WPI; 2002-280937/32.
XX	PT	New polypeptides for treating or preventing a disorder associated with them, in humans, e.g. cardiomyopathy, atherosclerosis or cancers.
XX	PS	Example 2; Page 234; 263pp; English.
XX	CC	The invention relates to an isolated polypeptide (NOVX) a mature form of NOVX, a NOVX variant (differing by no more than 15%), the nucleotide

CC encoding NOVX (or its complement, fragment or variant). NOVX is NOV1-14,  
 CC 15a, 15b, 16a, and 16b. The NOVX polypeptide, nucleic acid encoding it  
 CC and antibody against it, are useful for treating or preventing (e.g. by  
 CC gene therapy) a NOVX-associated disorder in humans, e.g. cardiomyopathy,  
 CC atherosclerosis, a disorder related to cell signal processing and  
 CC metabolic pathway modulation, diabetes or cancers. The NOVX polypeptide  
 CC and nucleic acids are also useful for determining the presence of  
 CC predisposition to the diseases. The NOVX nucleic acid and polypeptide are  
 CC especially useful in therapeutic or prophylactic applications for  
 CC disorders associated with aberrant NOVX expression or activity, e.g.  
 CC cancers (e.g. adenocarcinoma, lymphoma, prostate cancer or uterus  
 CC cancer), immune response, graft-versus-host disease, acquired  
 CC immunodeficiency syndrome (AIDS), asthma, Crohn's disease, hypertension,  
 CC congenital heart defects, multiple sclerosis, inflammation or Albrit  
 CC hereditary osteodystrophy and many other diseases listed in the  
 CC specification. The DNA encoding the protein is useful in gene therapy for  
 CC treating the conditions. This is also useful in detection assays,  
 CC chromosome mapping, tissue typing, diagnostic or prognostic assays, or  
 CC for developing a powerful assay system for functional analysis of various  
 CC human disorders, as well as in diagnostic applications. The present  
 CC sequence is a reverse transcriptase (RT)-PCR primer used to measure  
 CC tissue specific expression of mRNA encoding a NOVX protein  
 XX  
 SQ Sequence 22 BP; 4 A; 3 C; 8 G; 7 T; 0 U; 0 Other;  
 Query Match 100.0%; Score 22; DB 6; Length 22;  
 Best Local Similarity 100.0%; Pred. No. 0.79;  
 Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 1 TGAGAGAGTTCGGGTGCTCTA 22  
 DB 1 TGAGAGAGTTCGGGTGCTCTA 22  
 RESULT 2  
 ID ABK61465  
 AC ABK61465;  
 XX  
 XX 18-JUN-2002 (first entry)  
 DE Human CDNA encoding protein NOV13.  
 XX  
 KW Human; gene; ss; NOVX; gene therapy; cardiomyopathy; atherosclerosis;  
 KW cell signal processing disorder; metabolic pathway modulation disorder;  
 KW diabetes; cancer; adenocarcinoma; lymphoma; prostate cancer;  
 KW uterus cancer; immune response; graft-versus-host disease;  
 KW acquired immunodeficiency syndrome; AIDS; asthma; Crohn's disease;  
 KW hypertension; congenital heart defects; multiple sclerosis; inflammation;  
 KW Albrit hereditary osteodystrophy.  
 XX  
 OS Homo sapiens.  
 XX  
 XX WO200216599-A2.  
 XX  
 XX 28-FEB-2002.  
 XX  
 XX 27-AUG-2001; 2001WO-US026510.  
 XX  
 XX 25-AUG-2000; 2000US-0228191P.  
 XX 08-FEB-2001; 2001US-0267300P.  
 XX 20-FEB-2001; 2001US-0269961P.  
 XX 20-MAR-2001; 2001US-0277337P.  
 XX  
 XX (CURA-) CURAGEN CORP.  
 PA (CORT-) COR THERAPEUTICS INC.  
 XX  
 XX Burgess CE, Conley PB, Grosse WM, Hart M, Kekuda R, Shimkets RA;  
 PI Spytek KA, Szekeres ES, Tomlinson JE, Topper JN, Yang R;  
 XX WPI; 2002-280937/32.  
 DR P-PSDB; AAU91308.

XX New polypeptides for treating or preventing a disorder associated with  
 PT them, in humans, e.g. cardiomyopathy, atherosclerosis or cancers.  
 XX Claim 1; Page 98; 263pp; English.  
 XX  
 XX The invention relates to an isolated polypeptide (NOVX) a mature form of  
 CC NOVX, a NOVX variant (differing by no more than 15%), the nucleotide  
 CC encoding NOVX (or its complement, fragment or variant). NOVX is NOV1-14,  
 CC 15a, 15b, 16a, and 16b. The NOVX polypeptide, nucleic acid encoding it  
 CC and antibody against it, are useful for treating or preventing (e.g. by  
 CC gene therapy) a NOVX-associated disorder in humans, e.g. cardiomyopathy,  
 CC atherosclerosis, a disorder related to cell signal processing and  
 CC metabolic pathway modulation, diabetes or cancers. The NOVX polypeptide  
 CC and nucleic acids are also useful for determining the presence of  
 CC predisposition to the diseases. The NOVX nucleic acid and polypeptide are  
 CC especially useful in therapeutic or prophylactic applications for  
 CC disorders associated with aberrant NOVX expression or activity, e.g.  
 CC cancers (e.g. adenocarcinoma, lymphoma, prostate cancer or uterus  
 CC cancer), immune response, graft-versus-host disease, acquired  
 CC immunodeficiency syndrome (AIDS), asthma, Crohn's disease, hypertension,  
 CC congenital heart defects, multiple sclerosis, inflammation or Albrit  
 CC hereditary osteodystrophy and many other diseases listed in the  
 CC specification. The DNA encoding the protein is useful in gene therapy for  
 CC treating the conditions. This is also useful in detection assays,  
 CC chromosome mapping, tissue typing, diagnostic or prognostic assays, or  
 CC for developing a powerful assay system for functional analysis of various  
 CC human disorders, as well as in diagnostic applications. The present  
 CC sequence encodes a NOVX protein  
 XX  
 SQ Sequence 1183 BP; 251 A; 359 C; 333 G; 240 T; 0 U; 0 Other;  
 Query Match 100.0%; Score 22; DB 6; Length 1183;  
 Best Local Similarity 100.0%; Pred. No. 1.2;  
 Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 1 TGAGAGAGTTCGGGTGCTCTA 22  
 DB 224 TGAGAGAGTTCGGGTGCTCTA 245  
 RESULT 3  
 AAD43980  
 ID AAD43980 standard; cDNA; 2567 BP.  
 XX  
 AC AAD43980;  
 XX  
 DT 13-DEC-2002 (first entry)  
 XX  
 DE Human Src-Like Adapter Protein-2 (hSLAP-2) CDNA.  
 XX  
 KW Human; SH2/SH3-domain-containing adapter; Src-Like Adapter Protein-2;  
 KW SLAP-2; immune disorder; signal transduction; autoimmune disease; cancer;  
 KW neoplasm; inflammation disorder; rheumatoid arthritis; osteoarthritis;  
 KW psoriasis; rhinitis; allergy; inflammatory bowel disease; gene therapy;  
 KW Crohn's disease; systemic lupus erythematosus; tissue/organ rejection;  
 KW multiple sclerosis; asthma; acute respiratory distress syndrome;  
 KW pulmonary disorder; dermatological; neuroprotective; gene; ss.  
 XX  
 OS Homo sapiens.  
 XX  
 XX Key Location/Qualifiers  
 FT CDS 415..1200  
 FT /tag= a  
 FT /product= "Human SLAP-2"  
 XX  
 XX WO200242457-A1.  
 XX  
 XX 30-MAY-2002.  
 XX  
 XX 20-NOV-2001; 2001WO-US043367.  
 XX  
 XX 22-NOV-2000; 2000US-0252545P.

XX PA (BRIM ) BRISTOL-MYERS SQUIBB CO.

XX PI Chang H, Yang W, Wu Y, Whitney GS, Perez-Villar JJ, Kanner SB;

XX DR WPI; 2002-463632/49.

XX DR P-PSDB; AAE26357.

XX PT Novel substantially purified human SH2/SH3-domain-containing adapter

XX PT polypeptide, termed Src-like Adapter Protein-2, useful for therapeutic

XX PT intervention in immunological and inflammatory disorders and cancer.

XX PS Claim 2; Fig 1; 85pp; English.

XX CC The invention relates to a substantially purified human SH2/SH3-domain-

XX CC containing adapter polypeptide, termed Src-like Adapter Protein-2 (SLAP-

XX CC 2). The invention is useful for treating an immune disorder involving

XX CC hyperactivity of B- or T- lymphocytes in a mammal. The invention is

XX CC useful for screening for antagonists or inhibitors of the interaction of

XX CC hSLAP-2 with cellular signalling compounds, for diagnosing, treating or

XX CC preventing diseases or disorders associated with aberrant or uncontrolled

XX CC cellular signal transduction, for determining those cellular signalling

XX CC molecules which associate with hSLAP-2 and which provide critical signals

XX CC for cell activation, and as effectors in methods to affect T- cell

XX CC activation. The invention is useful in screening assays to identify and

XX CC detect candidate bioactive agents that modulate hSLAP-2 bioactivity, for

XX CC potential use to treat autoimmune diseases which may be caused by

XX CC hyperactivated B cells, as well as to treat diseases which may be caused

XX CC by hyperactivated T cells, in addition to other immune system related

XX CC conditions, diseases, or disorders. T-cell and B-cell neoplasms,

XX CC inflammatory disorders, diseases and conditions, rheumatoid arthritis,

XX CC osteoarthritis, psoriasis, rhinitis, inflammatory bowel disease (Crohn's

XX CC and ulcerative colitis), allergies, particularly those involving

XX CC hyperactivity of B-cells and T- cells, or other immune cells, such as

XX CC mast cells or eosinophils, autoimmune diseases such as systemic lupus

XX CC erythematosus and multiple sclerosis, pulmonary diseases including

XX CC asthma, acute respiratory distress syndrome, and chronic obstructive

XX CC pulmonary disorder, tissue/ organ rejection and cancer. The invention is

XX CC useful in gene therapy. The present sequence is human SLAP-2 cDNA

XX SQ Sequence 2567 BP; 611 A; 741 C; 666 G; 549 T; 0 U; 0 Other;

Query Match 100.0%; Score 22; DB 6; Length 2567;

Best Local Similarity 100.0%; Pred. No. 1.3;

Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 TGAGAGAGTTCGGGTGCTCA 22

DB 241 TGAGAGAGTTCGGGTGCTCA 262

RESULT 4

ABL28809

ID ABL28809 standard; DNA; 763 BP.

XX AC ABL28809;

XX DT 04-NOV-2002 (first entry)

XX DE Human ORF477 coding sequence.

XX KW Cytostatic; Cardiant; Anti-allergic; Immunosuppressive; Vulnary;

XX KW Antiinflammatory; gene therapy; human; ORFX; atherogenic; platelet;

XX KW human umbilical vein endothelial cell; HUVEC; atherosclerotic plaque;

XX KW cancer; cardiovascular disease; allergy; autoimmune disease;

XX KW wound healing; blood coagulation disorder; inflammatory disorder; ds.

XX OS Homo sapiens.

XX PN US2002082206-A1.

XX PD 27-JUN-2002.

PF 30-MAY-2001; 2001US-00867550.

XX PR 30-MAY-2000; 2000US-0208427P.

XX PA (LEAC/) LEACH M D.

XX PA (MEHR/) MEHRABAN F.

XX PA (CONL/) CONLEY P B.

XX PA (TOPP/) TOPPER J N.

XX PA (LAWD/) LAW D.

XX PI Leach MD, Mehraban F, Conley PB, Topper JN, Law D;

XX DR WPI; 2002-626554/67.

XX DR P-PSDB; ABP64107.

XX PT New polypeptide designated ORFX are present in human atherogenic cells

XX PT and are useful to prevent and treat ORFX-associated disorders including

XX PT cancer, allergy, wound healing or autoimmune, cardiovascular or

XX PT inflammatory disease.

XX PS Claim 2; SEQ ID NO 953; 78pp; English.

XX CC The present invention relates to novel human ORFX polypeptides and their

XX CC coding sequences (ABP63631-ABP64681 and ABQ98194-ABQ99267). The sequences

XX CC were discovered in human atherogenic cells, in particular in platelets

XX CC and human umbilical vein endothelial cells (HUVEC) and are expressed in

XX CC many other tissues as well. Atherogenic cells are cells which have the

XX CC potential to develop atherosclerotic plaques. The ORFX polypeptides and

XX CC nucleic acids are useful for treating or preventing a pathological

XX CC condition associated with an ORFX-associated disorder, e.g. cancer,

XX CC cardiovascular disease, allergy, autoimmune disease, wound healing, blood

XX CC coagulation disorders or inflammatory disorders. Note: The sequence data

XX CC for this patent did not form part of the printed specification, but was

XX CC obtained in electronic format directly from the USPTO web site at

XX CC seqdata.uspto.gov/sequence.html?docID=20020082205

XX SQ Sequence 763 BP; 176 A; 222 C; 218 G; 147 T; 0 U; 0 Other;

Query Match 92.7%; Score 20.4; DB 6; Length 763;

Best Local Similarity 95.5%; Pred. No. 6.7;

Matches 21; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 TGAGAGAGTTCGGGTGCTCA 22

DB 112 TGAGAGAGTTCGGGTGCTCA 133

RESULT 5

ABL28809

ID ABL28809 standard; DNA; 3319 BP.

XX AC ABL28809;

XX DT 26-MAR-2002 (first entry)

XX DE Drosophila melanogaster genomic polynucleotide SEQ ID NO 37900.

XX KW Drosophila; developmental biology; cell signalling; insecticide;

XX KW pharmaceutical; gene; ds.

XX OS Drosophila melanogaster.

XX PN WC200171042-A2.

XX PD 27-SEP-2001.

XX PF 23-MAR-2001; 2001WO-US009231.

XX PR 23-MAR-2000; 2000US-0191637P.

XX PR 11-JUL-2000; 2000US-00614150.

XX PA (PEKE ) PE CORP NY.

PI Venter JC, Adams M, Li PWD, Myers EW;  
XX WPI; 2001-656860/75.  
DR  
XX  
PT New isolated nucleic acid detection reagent for detecting 1000 or more  
PT genes from Drosophila and for elucidating cell signalling and cell-cell  
PT interactions.  
PT  
XX  
PS Claim 1; SEQ ID NO 37900; 21pp + Sequence Listing; English.  
XX  
CC The invention relates to an isolated nucleic acid detection reagent  
CC capable of detecting 1000 or more genes from Drosophila. The invention is  
CC useful in developmental biology and in elucidating cell signalling and  
CC cell-cell interactions in higher eukaryotes for the development of  
CC insecticides, therapeutics and pharmaceutical drugs. The invention  
CC discloses genomic DNA sequences (ABL16176-ABL30511), expressed DNA  
CC sequences (ABL01840-ABL16175) and the encoded proteins (ABB57737-  
CC ABB72072). The sequence data for this patent did not form part of the  
CC printed specification, but was obtained in electronic format directly  
CC from WIPO at ftp.wipo.int/pub/published\_pct\_sequences  
XX  
SQ Sequence 3319 BP; 699 A; 947 C; 992 G; 681 T; 0 U; 0 Other;  
Query Match 85.5%; Score 18.8; DB 4; Length 3319;  
Best Local Similarity 90.9%; Pred. No. 45;  
Matches 20; Conservative 0; Mismatches 2; Indels 0; Gaps 0;  
QY 1 TGAGAGAGTTCTGGGTGTCCTA 22  
Db 1729 TGAAGAGAGTTCTGGGTGTCCTA 1750  
RESULT 6  
ABL28808/c  
ID ABL28808 standard; DNA; 5782 BP.  
XX  
AC ABL28808;  
XX  
DT 26-MAR-2002 (first entry)  
XX  
DE Drosophila melanogaster genomic polynucleotide SEQ ID NO 37897.  
XX  
KW Drosophila; developmental biology; cell signalling; insecticide;  
KW pharmaceutical; gene; ds.  
XX  
OS Drosophila melanogaster.  
XX  
PN WO200171042-A2.  
XX  
PD 27-SEP-2001.  
XX  
PF 23-MAR-2001; 2001WO-US009231.  
XX  
PR 23-MAR-2000; 2000US-0191637P.  
PR 11-JUL-2000; 2000US-00614150.  
XX  
PA (PEKE ) PE CORP NY.  
XX  
PI Venter JC, Adams M, Li PWD, Myers EW;  
XX WPI; 2001-656860/75.  
DR  
XX  
PT New isolated nucleic acid detection reagent for detecting 1000 or more  
PT genes from Drosophila and for elucidating cell signalling and cell-cell  
PT interactions.  
PT  
XX  
PS Claim 1; SEQ ID NO 37897; 21pp + Sequence Listing; English.  
XX  
CC The invention relates to an isolated nucleic acid detection reagent  
CC capable of detecting 1000 or more genes from Drosophila. The invention is  
CC useful in developmental biology and in elucidating cell signalling and  
CC cell-cell interactions in higher eukaryotes for the development of  
CC insecticides, therapeutics and pharmaceutical drugs. The invention

CC discloses genomic DNA sequences (ABL16176-ABL30511), expressed DNA  
CC sequences (ABL01840-ABL16175) and the encoded proteins (ABB57737-  
CC ABB72072). The sequence data for this patent did not form part of the  
CC printed specification, but was obtained in electronic format directly  
CC from WIPO at ftp.wipo.int/pub/published\_pct\_sequences  
XX  
SQ Sequence 5782 BP; 1326 A; 1604 C; 1561 G; 1291 T; 0 U; 0 Other;  
Query Match 85.5%; Score 18.8; DB 4; Length 5782;  
Best Local Similarity 90.9%; Pred. No. 48;  
Matches 20; Conservative 0; Mismatches 2; Indels 0; Gaps 0;  
QY 1 TGAGAGAGTTCTGGGTGTCCTA 22  
Db 2789 TGAAGAGAGTTCTGGGTGTCCTA 2768  
RESULT 7  
ABX70675  
ID ABX70675 standard; cDNA; 2442 BP.  
XX  
AC ABX70675;  
XX  
DT 04-MAR-2003 (first entry)  
XX  
DE Human cDNA encoding NOV26.  
XX  
KW Human; ss; gene; NOVX; metabolic disorder; diabetes; cardiomyopathy;  
KW obesity; infectious disease; anorexia; neurodegenerative disorder;  
KW Alzheimer's disease; Parkinson's disease; immune disorder;  
KW haematopoietic disorder; dyslipidaemia; metabolic disturbance;  
KW metabolic syndrome X; wasting disorder; cancer; gene therapy; SNP;  
KW single nucleotide polymorphism.  
XX  
OS Homo sapiens.  
XX  
PH Key Location/Qualifiers  
FT variation replace(72,C)  
FT /\*tag= a  
FT /\*standard\_name= "Single nucleotide polymorphism"  
FT variation replace(78,C)  
FT /\*tag= b  
FT /\*standard\_name= "Single nucleotide polymorphism"  
FT variation replace(201,C)  
FT /\*tag= c  
FT /\*standard\_name= "Single nucleotide polymorphism"  
FT variation replace(412,T)  
FT /\*tag= d  
FT /\*standard\_name= "Single nucleotide polymorphism"  
FT variation replace(2253,G)  
FT /\*tag= e  
FT /\*standard\_name= "Single nucleotide polymorphism"  
XX  
PN WO200281518-A2.  
XX  
PD 17-OCT-2002.  
XX  
PF 21-FEB-2002; 2002WO-US005374.  
XX  
PR 21-FEB-2001; 2001US-0270220P.  
PR 21-FEB-2001; 2001US-0270523P.  
PR 23-FEB-2001; 2001US-0270797P.  
PR 23-FEB-2001; 2001US-0270810P.  
PR 08-MAR-2001; 2001US-0274295P.  
PR 16-MAR-2001; 2001US-0276400P.  
PR 16-MAR-2001; 2001US-0276677P.  
PR 26-MAR-2001; 2001US-0278796P.  
PR 04-APR-2001; 2001US-0281521P.  
PR 25-APR-2001; 2001US-0286548P.  
PR 17-MAY-2001; 2001US-0291765P.  
PR 10-AUG-2001; 2001US-0311595P.  
PR 13-AUG-2001; 2001US-0311980P.  
PR 10-SEP-2001; 2001US-0318526P.

PR 17-SEP-2001; 2001US-0322712P.  
 PR 18-OCT-2001; 2001US-0330307P.  
 PA (CURA-) CURAGEN CORP.  
 XX Pena CEA, Shimkets RA, Li L, Shenoy SG, Kekuda R, Spytek KA;  
 PI Vernet CAM, Malyankar UL, Guo X, Gusev VY, Casman SJ, Boldog FL;  
 PI Furtak K, Tchernev VT, Patturajan M, Gangolli EA, Radigar M, Liu X;  
 PI Baumgartner JC, Gerlach VL, Spaderna SK, Zerhusen BD;  
 XX WPI; 2003-045859/04.  
 DR P-PSDB; ABUS2620.  
 XX  
 FT New isolated NOVX polypeptide useful for treating cardiomyopathy,  
 FT atherosclerosis, metabolic disorders, diabetes, obesity, infectious  
 PT disease, anorexia, neurodegenerative disorders, Alzheimer's disease and  
 PT cancer.  
 XX  
 PS Claim 9; Page 210-211; 479pp; English.  
 XX  
 CC The invention relates to an isolated polypeptide termed NOVX (NOV1, 2a,  
 CC 2b, 3a, 3b, 4a, 4b, 5, 6, 7a-c, 8a-e, 9a-b, 10, 11, 12a-c, 13, 14, 15,  
 CC 16a-d, 17a-b, 18, 19, 20a-b, 21-30) appearing as ABUS2578-ABUS2624) a  
 CC variant of NOVX, a mature form of NOVX, and a variant of the mature form  
 CC of NOVX. Also included are a nucleic acid molecule (NOVX NA) encoding  
 CC NOVX, or a fragment or complement of NOVX NA, a vector comprising NOVX  
 CC NA, a cell comprising the vector, an anti-NOVX antibody (ab), determining  
 CC the presence or amount of NOVX or NOVX NA in a sample, and identifying an  
 CC agent that binds or modulates the expression or activity of NOVX. NOVX,  
 CC NOVX NA or ab is useful for treating or preventing a NOVX-associated  
 CC disorder in a subject, preferably human. Ab is useful for determining the  
 CC presence or amount of NOVX in a sample. NOVX NA or ab is useful for identifying an  
 CC agent that binds to NOVX. NOVX, NOVX NA or ab is useful for treating  
 CC metabolic disorders, diabetes, cardiomyopathy, obesity, infectious  
 CC disease, anorexia, neurodegenerative disorders, Alzheimer's disease,  
 CC Parkinson's disease, immune disorders, haematopoietic disorders, and  
 CC various dyslipidaemias, metabolic disturbances associated with obesity,  
 CC the metabolic syndrome X and wasting disorders associated with chronic  
 CC diseases, various cancers, endocrine, connective tissue, blood, vascular,  
 CC skin, renal, bone, brain, muscle disorders, or bacterial, fungal,  
 CC protozoal or viral infections. NOVX, NOVX NA or ab is useful in screening  
 CC assays, detection assays, predictive medicine, and in methods of  
 CC treatment. NOVX is useful as immunogen, to screen for potential  
 CC ant/agonist compounds, and as bait protein in a two-hybrid or three-  
 CC hybrid assay. NOVX NA is useful in gene therapy, to express NOVX, to  
 CC detect NOVX mRNA or a genetic lesion in a NOVX gene, and to modulate NOVX  
 CC activity. The cell is useful for producing non-human transgenic animals.  
 CC Ab is useful for isolating, and purifying NOVX and to monitor protein  
 CC levels in tissue as part of a clinical testing procedure. The present  
 CC sequence encodes a NOVX protein  
 XX  
 SQ Sequence 2442 BP; 645 A; 631 C; 643 G; 523 T; 0 U; 0 Other;  
 Query Match 79.1%; Score 17.4; DB 7; Length 2442;  
 Best Local Similarity 94.7%; Pred. No. 2e-02;  
 Matches 18; Conservative 0; Mismatches 1; Indels 0; Gaps 0;  
 OY 4 GAGAGTTCGGTGCTCTA 22  
 DB 1116 GAGAGTTCGGATGCTCTA 1134  
 RESULT 8  
 ID AAD30568  
 ID AAD30568 standard; cDNA; 2789 BP.  
 AC AAD30568;  
 XX  
 DT 21-MAY-2002 (first entry)  
 XX Human protease, PRTS-1 cDNA.  
 DE Human; protease; PRTS-1; enzyme; gastritis; cirrhosis; Crohn's disease;  
 KW Human; protease; PRTS-1; enzyme; gastritis; cirrhosis; Crohn's disease;

KW gastrointestinal disorder; autoimmune; inflammatory; cell proliferative;  
 KW cardiovascular; developmental; epithelial; neurological; reproductive;  
 KW AIDS; Acquired Immune Deficiency Syndrome; allergy; rheumatoid arthritis;  
 KW anaemia; asthma; atherosclerosis; hypertension; myocardial infarction;  
 KW hepatitis; cancer; psoriasis; Cushing's syndrome; hypothyroidism; eczema;  
 KW epilepsy; Alzheimer's disease; Huntington's disease; Parkinson's disease;  
 XX Pick's disease; infertility; vitiligo; drug screening; gene therapy; ss.  
 OS Homo sapiens.  
 XX  
 FH Key Location/Qualifiers  
 FT CDS 303..2606  
 FT /tag= a  
 FT /product= "Human PRTS-1 protein"  
 FT sig\_peptide 303..347  
 FT /tag= b  
 FT mat\_peptide 348..2603  
 FT /tag= c  
 FT /product= "Mature PRTS-1 protein"  
 XX  
 XX WO200208396-A2.  
 XX  
 PD 31-JAN-2002.  
 XX  
 PF 17-JUL-2001; 2001WO-US022397.  
 XX  
 PR 21-JUL-2000; 2000US-0220063P.  
 PR 28-JUL-2000; 2000US-0221680P.  
 PR 04-AUG-2000; 2000US-0223544P.  
 PR 11-AUG-2000; 2000US-0224717P.  
 PR 16-AUG-2000; 2000US-0225988P.  
 PR 23-AUG-2000; 2000US-0227568P.  
 XX (INCV-) INCVTE GENOMICS INC.  
 XX  
 PA Deleage AM, Gandhi AR, Hafalia AJA, Lu DAM, Patterson C;  
 PI Tribouley CM, Das D, Kallick DA, Nguyen DB, Lee EA, Khan FA, Yue H;  
 PI Au-Young J, Griffin JA, Policky JL, Ramkumar J, Yang J;  
 PI Thangavelu K, Ding L, Kearney L, Baughn MR, Borowsky ML;  
 PI Sanjanwala MS, Yao MG, Burford N, Wallia NK, Lal P, Lee S, Todd S;  
 PI Lo TP, Tang YT, Elliott VS, Azimzai Y, Lu Y;  
 XX WPI; 2002-206082/26.  
 DR P-PSDB; AAE19164.  
 XX  
 XX New human protease polypeptide, useful in diagnosis, prevention and  
 PT treatment of gastrointestinal, cardiovascular, autoimmune/inflammatory,  
 PT cell proliferative, developmental, epithelial and neurological disorders.  
 XX  
 PS Claim 5; Page 164-165; 182pp; English.  
 XX  
 CC The invention relates to an isolated human protease polypeptide (PRTS).  
 CC PRTS protein and DNA are useful for diagnosing, treating and preventing  
 CC gastrointestinal disorders (gastritis, cirrhosis, Crohn's disease),  
 CC autoimmune/inflammatory disorders (AIDS, allergy, rheumatoid arthritis,  
 CC anaemia, asthma), cardiovascular disorder (atherosclerosis, hypertension,  
 CC myocardial infarction), cell proliferative disorders (hepatitis, cancer,  
 CC psoriasis), developmental disorders (Cushing's syndrome, hypothyroidism),  
 CC epithelial disorder (vitiligo, keloid, eczema), neurological disorders  
 CC (epilepsy, Alzheimer's disease, Pick's disease, Huntington's disease,  
 CC Parkinson's disease), and reproductive disorders (infertility). PRTS  
 CC protein is useful in a number of drug screening techniques and to analyse  
 CC the proteome of a tissue or cell type. PRTS DNA is useful for creating  
 CC knockin humanised animals or transgenic animals to model human diseases,  
 CC in somatic or germline gene therapy and in microarrays utilising fluids  
 CC or tissues from patients to detect altered PKIN expression. The present  
 CC sequence is human PRTS-1 cDNA  
 XX  
 SQ Sequence 2789 BP; 715 A; 745 C; 698 G; 631 T; 0 U; 0 Other;  
 Query Match 79.1%; Score 17.4; DB 6; Length 2789;  
 Best Local Similarity 94.7%; Pred. No. 2.1e+02;  
 Matches 18; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 4 GAGAGTCTGGGTGCTCTA 22  
 |||||  
 Db 1342 GAGAGTCTGGGTGCTCTA 1360

RESULT 9  
 AAH57383  
 ID AAH57383 standard; cDNA; 3076 BP.  
 XX  
 AC AAH57383;  
 XX  
 DT 10-SEP-2001 (first entry)  
 XX  
 DE Human skeletal muscle cell specific cDNA sequence SEQ ID NO:223.  
 XX  
 KW Human; tissue specific; diagnosis; brain; heart; skeletal muscle; lung;  
 KW liver; uterus; ovary; stomach; intestine; kidney; pancreas; se;  
 KW metabolic disease; developmental disease; cystostatic; immunomodulatory;  
 KW neuroprotective; gene therapy; cancer; immunopathology; neuropathology.  
 XX  
 OS Homo sapiens.  
 XX  
 PN WO200132927-A2.  
 XX  
 PD 10-MAY-2001.  
 XX  
 PF 02-NOV-2000; 2000WO-US030396.  
 XX  
 PR 04-NOV-1999; 99US-0163508P.  
 XX  
 PA (INCY-) INCYTE GENOMICS INC.  
 XX  
 PI Sornasse T, Seilhamer JU, Watson GA;  
 XX  
 DR WPI; 2001-291057/30.  
 XX  
 DT New cell and tissue specific polynucleotides useful for diagnosis,  
 PT prognosis or monitoring of treatments for disorders where the gene is  
 PT associated with a cancer, immunopathology or neuropathology.  
 XX  
 PS Claim 1; Page 153-154; 327pp; English.

XX  
 CC AAH57161 to AAH57576 represent cell and tissue specific polynucleotide  
 CC sequences (I). (I) can have cytosstatic, immunomodulatory and  
 CC neuroprotective activities, and can be used in gene therapy. (I) and  
 CC proteins (II) encoded by then are used in high throughput screening  
 CC assays to select DNA molecules, RNA molecules, peptide nucleic acids,  
 CC mimetics, peptides, proteins, agonists, antagonists, antibodies or their  
 CC fragments, immunoglobulins, inhibitors, drug compounds and pharmaceutical  
 CC agents. Expression of (I) in a sample indicates the differentiation of  
 CC embryonic stem cells into a tissue selected from brain, heart, kidney,  
 CC liver, lung, skeletal muscle or pancreatic tissues. (I) and (II) are used  
 CC to produce an expression profile that defines a metabolic or  
 CC developmental process, treatment, condition, disease or disorder. The  
 CC gene profile can be used for diagnosis, prognosis or monitoring of  
 CC treatments and for investigating a predisposition to a disorder where the  
 CC gene is associated with a cancer, immunopathology or neuropathology  
 XX  
 SQ Sequence 3076 BP; 800 A; 814 C; 792 G; 670 T; 0 U; 0 Other;  
 Query Match 79.1%; Score 17.4; DB 4; Length 3076;  
 Best Local Similarity 94.7%; Pred. No. 2.1e+02;  
 Matches 18; Conservative 0; Mismatches 1; Indels 0; Gaps 0;  
 QY 4 GAGAGTCTGGGTGCTCTA 22  
 |||||  
 Db 1244 GAGAGTCTGGGTGCTCTA 1262

RESULT 10  
 AAS90933  
 ID AAS90933 standard; cDNA; 3210 BP.

XX AAS90933;  
 AC  
 XX 13-FEB-2002 (first entry)  
 DT  
 XX  
 DE DNA encoding novel human diagnostic protein #26737.  
 XX  
 KW Human; chromosome mapping; gene mapping; gene therapy; forensic;  
 KW food supplement; medical imaging; diagnostic; genetic disorder; ss.  
 XX  
 OS Homo sapiens.  
 XX  
 PN WO200175067-A2.  
 XX  
 PD 11-OCT-2001.  
 XX  
 PF 30-MAR-2001; 2001WO-US008631.  
 XX  
 PR 31-MAR-2000; 2000US-00540217.  
 PR 23-AUG-2000; 2000US-00649167.  
 XX  
 PA (HYSE-) HYSEQ INC.  
 XX  
 PI Drmanac RT, Liu C, Tang YT;  
 XX  
 DR WPI; 2001-639362/73.  
 DR P-PSDB; ABG26746.  
 XX  
 PT New isolated polynucleotide and encoded polypeptides, useful in  
 PT diagnostics, forensics, gene mapping, identification of mutations  
 PT responsible for genetic disorders or other traits and to assess  
 PT biodiversity.  
 XX  
 PS Claim 1; SEQ ID NO 26737; 103pp; English.

XX  
 CC The invention relates to isolated polynucleotide (I) and polypeptide (II)  
 CC sequences (I) is useful as hybridisation probes, polymerase chain  
 CC reaction (PCR) primers, oligomers, and for chromosome and gene mapping,  
 CC and in recombinant production of (II). The polynucleotides are also used  
 CC in diagnostics as expressed sequence tags for identifying expressed  
 CC genes. (I) is useful in gene therapy techniques to restore normal  
 CC activity of (II) or to treat disease states involving (II). (II) is  
 CC useful for generating antibodies against it, detecting or quantitating a  
 CC polypeptide in tissue, as molecular weight markers and as a food  
 CC supplement. (II) and its binding partners are useful in medical imaging  
 CC of sites expressing (II). (I) and (II) are useful for treating disorders  
 CC involving aberrant protein expression or biological activity. The  
 CC polypeptide and polynucleotide sequences have application in  
 CC diagnostics, forensics, gene mapping, identification of mutations  
 CC responsible for genetic disorders or other traits to assess biodiversity  
 CC and to produce other types of data and products dependent on DNA and  
 CC amino acid sequences. AAS64197-AAS94564 represent novel human diagnostic  
 CC coding sequences of the invention. Note: The sequence data for this  
 CC patent did not appear in the printed specification, but was obtained in  
 CC electronic format directly from WIPO at  
 CC ftp.wipo.int/pub/published\_pct\_sequences  
 XX

SQ Sequence 3210 BP; 821 A; 839 C; 861 G; 689 T; 0 U; 0 Other;  
 Query Match 79.1%; Score 17.4; DB 5; Length 3210;  
 Best Local Similarity 94.7%; Pred. No. 2.1e+02;  
 Matches 18; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 4 GAGAGTCTGGGTGCTCTA 22  
 |||||  
 Db 1784 GAGAGTCTGGGTGCTCTA 1802

RESULT 11  
 ACH03882  
 ID ACH03882 standard; cDNA; 3327 BP.  
 XX  
 AC ACH03882;

```

XX DT 26-SEP-2003 (first entry)
XX DE Human cDNA differentially expressed in lung cancer #87.
XX KW Gene therapy; emphysema; ss; gene; chronic obstructive pulmonary disease;
XX KW respiratory disorder; lung cancer; asthma; human.
XX OS Homo sapiens.
XX PN US2003065157-A1.
XX PD 03-APR-2003.
XX XX 04-APR-2002; 2002US-00116802.
XX PF 04-APR-2001; 2001US-0281593P.
XX PR (LASE/) LASEK A W.
XX PA Lasek AW;
XX PI WPI; 2003-540803/51.
XX DR New combination comprising cDNAs that are differentially expressed in
XX PT respiratory disorders, useful for diagnosing or treating respiratory
XX PT disorders e.g., lung cancer, chronic obstructive pulmonary disease,
XX PT emphysema or asthma.
XX PS Claim 1; Page; 39pp; English.
XX CC The invention relates to a combination comprising cDNAs or their
XX CC complements that are differentially expressed in respiratory disorder.
XX CC The combination is useful for preparing a composition for diagnosing or
XX CC treating respiratory disorders e.g. lung cancer, chronic obstructive
XX CC pulmonary disease, emphysema or asthma. The present sequence represents
XX CC human cDNA differentially expressed during lung cancer
XX SQ Sequence 3327 BP; 863 A; 880 C; 834 G; 750 T; 0 U; 0 Other;
Query Match 79.1%; Score 17.4; DB 8; Length 3327;
Best Local Similarity 94.7%; Pred. No. 2.1e+02;
Matches 18; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
QY 4 GAGAGTTCTGGTGTCCTA 22
Db 1486 GAGAGTTCTGGATGTCCTA 1504
RESULT 12
AAT32455
ID AAT32455 standard; cDNA; 5149 BP.
AC AAT32455;
XX 02-DEC-1996 (first entry)
XX Calpain large subunit 1 coding sequence.
XX Calpain; subunit; calcium; protease; mutation; treatment; detection;
XX KW identification; diagnosis; limb girdle muscular dystrophy; LGMD2;
XX KW calcium activated neutral protease; CANP; ss.
XX OS Homo sapiens.
XX PH Key Location/Qualifiers
XX FT misc_feature 109..359
XX FT /tag= a
XX FT /label= Alu sequence.
XX FT CAAT_signal 712..716
XX FT /tag= b
XX FT /note= "Putative CAAT signal."
XX FT CAAT_signal 759..764

```

```

FT CAAT_signal
FT /tag= C
FT /note= "Putative CAAT signal."
FT 979..983
FT /tag= d
FT TATA_signal
FT /note= "Putative CAAT signal."
FT 1270..1275
FT /tag= e
FT /note= "Putative TATA signal."
FT 1303..3768
FT /tag= f
FT /product= "Calpain large subunit 1."
FT 4288..4293
FT /tag= g
XX WO9616175-A2.
XX PN 30-MAY-1996.
XX PD 21-NOV-1995; 95WO-EP004575.
XX PF 22-NOV-1994; 94EP-00402668.
XX PR (ASPR-) ASSOC FR CONTRE MYOPATHIES.
XX PA Beckmann J, Richard I;
XX PI WPI; 1996-268611/27.
XX DR P-PSDB; AAR99579.
XX XX Human novel Calpain large subunit 1 gene encoding a calcium dependent
XX PT protease - used to develop prods. for the diagnosis and treatment of limb
XX PT -girdle muscular dystrophy 2 disease.
XX PS Claim 1; Fig 2; 66pp; English.
XX CC The calpain large subunit 1 gene located on chromosome 15 codes for a
XX CC calcium activated neutral protease (CANP3) belonging to the calpain
XX CC family. Mutations in the gene induce limb-girdle muscular dystrophy
XX CC (LGMD) 2 disease. The gene, and fragments of it, can be used in the
XX CC prevention, treatment, diagnosis and detection of a predisposition to
XX CC LGMD2 disease
XX XX Sequence 5149 BP; 1343 A; 1312 C; 1179 G; 1314 T; 0 U; 1 Other;
Query Match 79.1%; Score 17.4; DB 2; Length 5149;
Best Local Similarity 94.7%; Pred. No. 2.2e+02;
Matches 18; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
QY 4 GAGAGTTCTGGTGTCCTA 22
Db 2486 GAGAGTTCTGGATGTCCTA 2504
RESULT 13
AAK67913/c
ID AAK67913 standard; DNA; 355 BP.
XX AAK67913;
XX 06-NOV-2001 (first entry)
XX Human immune/haematopoietic antigen genomic sequence SEQ ID NO:22725.
XX KW Human; immune; haematopoietic; immune/haematopoietic antigen; cancer;
XX KW cytosolic; gene therapy; vaccine; metastasis; ds.
XX OS Homo sapiens.
XX PN WO200157182-A2.
XX PD 09-AUG-2001.
XX PF 17-JAN-2001; 2001WO-US001354.

```



XX 31-JAN-2000; 2000US-0179065P.  
 PR 04-FEB-2000; 2000US-0180282P.  
 PR 24-FEB-2000; 2000US-0184664P.  
 PR 02-MAR-2000; 2000US-0186350P.  
 PR 16-MAR-2000; 2000US-0189874P.  
 PR 17-MAR-2000; 2000US-0190076P.  
 PR 18-APR-2000; 2000US-0198123P.  
 PR 19-MAY-2000; 2000US-0205515P.  
 PR 07-JUN-2000; 2000US-0209467P.  
 PR 28-JUN-2000; 2000US-0214886P.  
 PR 30-JUN-2000; 2000US-0215135P.  
 PR 07-JUL-2000; 2000US-0216647P.  
 PR 07-JUL-2000; 2000US-0216880P.  
 PR 11-JUL-2000; 2000US-0217487P.  
 PR 11-JUL-2000; 2000US-0217496P.  
 PR 14-JUL-2000; 2000US-0218290P.  
 PR 26-JUL-2000; 2000US-0220963P.  
 PR 26-JUL-2000; 2000US-0220964P.  
 PR 14-AUG-2000; 2000US-02224518P.  
 PR 14-AUG-2000; 2000US-02224519P.  
 PR 14-AUG-2000; 2000US-0225213P.  
 PR 14-AUG-2000; 2000US-0225214P.  
 PR 14-AUG-2000; 2000US-0225266P.  
 PR 14-AUG-2000; 2000US-0225267P.  
 PR 14-AUG-2000; 2000US-0225268P.  
 PR 14-AUG-2000; 2000US-0225270P.  
 PR 14-AUG-2000; 2000US-0225447P.  
 PR 14-AUG-2000; 2000US-0225757P.  
 PR 14-AUG-2000; 2000US-0225758P.  
 PR 14-AUG-2000; 2000US-0225759P.  
 PR 18-AUG-2000; 2000US-0226279P.  
 PR 22-AUG-2000; 2000US-0226581P.  
 PR 22-AUG-2000; 2000US-0226582P.  
 PR 22-AUG-2000; 2000US-0227182P.  
 PR 23-AUG-2000; 2000US-0227009P.  
 PR 30-AUG-2000; 2000US-0228924P.  
 PR 01-SEP-2000; 2000US-02329287P.  
 PR 01-SEP-2000; 2000US-0232943P.  
 PR 01-SEP-2000; 2000US-0233443P.  
 PR 01-SEP-2000; 2000US-0233444P.  
 PR 05-SEP-2000; 2000US-0233945P.  
 PR 05-SEP-2000; 2000US-0233950P.  
 PR 06-SEP-2000; 2000US-0234113P.  
 PR 06-SEP-2000; 2000US-02340437P.  
 PR 06-SEP-2000; 2000US-02340438P.  
 PR 08-SEP-2000; 2000US-0231243P.  
 PR 08-SEP-2000; 2000US-0231243P.  
 PR 08-SEP-2000; 2000US-0231244P.  
 PR 08-SEP-2000; 2000US-0231413P.  
 PR 08-SEP-2000; 2000US-0231414P.  
 PR 08-SEP-2000; 2000US-0232080P.  
 PR 08-SEP-2000; 2000US-0232081P.  
 PR 12-SEP-2000; 2000US-0231968P.  
 PR 14-SEP-2000; 2000US-0233397P.  
 PR 14-SEP-2000; 2000US-0233398P.  
 PR 14-SEP-2000; 2000US-0233399P.  
 PR 14-SEP-2000; 2000US-0233400P.  
 PR 14-SEP-2000; 2000US-0233401P.  
 PR 14-SEP-2000; 2000US-0233063P.  
 PR 14-SEP-2000; 2000US-0233064P.  
 PR 14-SEP-2000; 2000US-0233065P.  
 PR 21-SEP-2000; 2000US-0234223P.  
 PR 21-SEP-2000; 2000US-0234224P.  
 PR 25-SEP-2000; 2000US-0234997P.  
 PR 25-SEP-2000; 2000US-0234998P.  
 PR 26-SEP-2000; 2000US-0234844P.  
 PR 27-SEP-2000; 2000US-0235834P.  
 PR 27-SEP-2000; 2000US-0235835P.  
 PR 29-SEP-2000; 2000US-0236327P.  
 PR 29-SEP-2000; 2000US-0236367P.  
 PR 29-SEP-2000; 2000US-0236368P.  
 PR 29-SEP-2000; 2000US-0236369P.  
 PR 29-SEP-2000; 2000US-0236370P.

PR 02-OCT-2000; 2000US-0236802P.  
 PR 02-OCT-2000; 2000US-0237037P.  
 PR 02-OCT-2000; 2000US-0237038P.  
 PR 02-OCT-2000; 2000US-0237039P.  
 PR 13-OCT-2000; 2000US-0237040P.  
 PR 13-OCT-2000; 2000US-0239935P.  
 PR 20-OCT-2000; 2000US-0239937P.  
 PR 20-OCT-2000; 2000US-0240960P.  
 PR 20-OCT-2000; 2000US-0241211P.  
 PR 20-OCT-2000; 2000US-0241785P.  
 PR 20-OCT-2000; 2000US-0241786P.  
 PR 20-OCT-2000; 2000US-0241787P.  
 PR 20-OCT-2000; 2000US-0241808P.  
 PR 20-OCT-2000; 2000US-0241809P.  
 PR 01-NOV-2000; 2000US-0241826P.  
 PR 01-NOV-2000; 2000US-0244617P.  
 PR 08-NOV-2000; 2000US-0246474P.  
 PR 08-NOV-2000; 2000US-0246475P.  
 PR 08-NOV-2000; 2000US-0246476P.  
 PR 08-NOV-2000; 2000US-0246477P.  
 PR 08-NOV-2000; 2000US-0246478P.  
 PR 08-NOV-2000; 2000US-0246523P.  
 PR 08-NOV-2000; 2000US-0246524P.  
 PR 08-NOV-2000; 2000US-0246525P.  
 PR 08-NOV-2000; 2000US-0246526P.  
 PR 08-NOV-2000; 2000US-0246527P.  
 PR 08-NOV-2000; 2000US-0246528P.  
 PR 08-NOV-2000; 2000US-0246532P.  
 PR 08-NOV-2000; 2000US-0246609P.  
 PR 08-NOV-2000; 2000US-0246610P.  
 PR 08-NOV-2000; 2000US-0246611P.  
 PR 08-NOV-2000; 2000US-0246613P.  
 PR 17-NOV-2000; 2000US-0249207P.  
 PR 17-NOV-2000; 2000US-0249208P.  
 PR 17-NOV-2000; 2000US-0249209P.  
 PR 17-NOV-2000; 2000US-0249210P.  
 PR 17-NOV-2000; 2000US-0249211P.  
 PR 17-NOV-2000; 2000US-0249212P.  
 PR 17-NOV-2000; 2000US-0249213P.  
 PR 17-NOV-2000; 2000US-0249214P.  
 PR 17-NOV-2000; 2000US-0249215P.  
 PR 17-NOV-2000; 2000US-0249216P.  
 PR 17-NOV-2000; 2000US-0249217P.  
 PR 17-NOV-2000; 2000US-0249218P.  
 PR 17-NOV-2000; 2000US-0249244P.  
 PR 17-NOV-2000; 2000US-0249245P.  
 PR 17-NOV-2000; 2000US-0249264P.  
 PR 17-NOV-2000; 2000US-0249265P.  
 PR 17-NOV-2000; 2000US-0249297P.  
 PR 17-NOV-2000; 2000US-0249299P.  
 PR 01-DEC-2000; 2000US-0250160P.  
 PR 01-DEC-2000; 2000US-0250391P.  
 PR 05-DEC-2000; 2000US-0251030P.  
 PR 05-DEC-2000; 2000US-0251988P.  
 PR 06-DEC-2000; 2000US-0256719P.  
 PR 06-DEC-2000; 2000US-0251479P.  
 PR 08-DEC-2000; 2000US-0251856P.  
 PR 08-DEC-2000; 2000US-0251868P.  
 PR 08-DEC-2000; 2000US-0251869P.  
 PR 08-DEC-2000; 2000US-0251989P.  
 PR 11-DEC-2000; 2000US-0251990P.  
 PR 11-DEC-2000; 2000US-0254097P.  
 PR 05-JAN-2001; 2001US-0259678P.

(HUMA-) HUMAN GENOME SCI INC.

Rosen CA, Barash SC, Ruben SM;

WPI; 2001-483426/52.

Nucleic acids encoding human immune/hematopoietic antigen polypeptides,  
 useful for preventing, diagnosing and/or treating cancers and metastasis.

XX Disclosure; SEQ ID NO 22725; 3071pp + Sequence Listing; English.  
 PS  
 CC AAK54951 to AAK64702 encode the human immune/haematopoietic antigen (I)  
 CC amino acid sequences given in AAK64702 to AAK64702. (I) have cytostatic  
 CC activity, and can be used in gene therapy and vaccine production. (I)  
 CC proteins and polynucleotides may be used in the prevention, diagnosis and  
 CC treatment of diseases associated with inappropriate (I) expression. For  
 CC example, they may be used to treat disorders associated with decreased  
 CC expression by rectifying mutations or deletions in a patient's genome  
 CC that affect the activity of (I) by expressing inactive proteins or to  
 CC supplement the patient's own production of (I). Additionally, (I)  
 CC polynucleotides may be used to produce the secreted (I), by inserting the  
 CC nucleic acids into a host cell and culturing the cell to express the  
 CC protein. (I) proteins and polynucleotides may be used to prevent,  
 CC diagnose and treat immune/haematopoietic-related diseases, especially  
 CC cancers and cancer metastases of haematopoietic-derived cells. AAK64703  
 CC to AAK64794 represent human immune/haematopoietic antigen genomic  
 CC sequences from the present invention. AAK54942 to AAK54950 and AAK62169  
 CC represent sequences used in the exemplification of the present invention  
 XX  
 SQ Sequence 355 BP; 121 A; 77 C; 72 G; 85 T; 0 U; 0 Other;

Query Match 78.2%; Score 17.2; DB 4; Length 355;  
 Best Local Similarity 86.4%; Pred. No. 2.1e+02;  
 Matches 19; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 TGAGAGAGTCTCGGTCTCTA 22  
 |||||  
 Db 146 TGAGTGTGTTCTGGCTCTCTA 125

RESULT 14  
 AAK67912/c  
 ID AAK67912 standard; DNA; 356 BP.  
 AC AAK67912;  
 XX  
 DT 06-NOV-2001 (first entry)  
 XX  
 DE Human immune/haematopoietic antigen genomic sequence SEQ ID NO:22724.  
 XX  
 KW Human; immune; haematopoietic; immune/haematopoietic antigen; cancer;  
 KM cytostatic; gene therapy; vaccine; metastasis; ds.  
 XX  
 OS Homo sapiens.  
 XX  
 PN WO200157182-A2.  
 XX  
 PD 09-AUG-2001.  
 XX  
 PF 17-JAN-2001; 2001WO-US001354.  
 XX  
 PR 31-JAN-2000; 2000US-0179065P.  
 PR 04-FEB-2000; 2000US-0180628P.  
 PR 24-FEB-2000; 2000US-0184664P.  
 PR 02-MAR-2000; 2000US-0186350P.  
 PR 16-MAR-2000; 2000US-0189874P.  
 PR 17-MAR-2000; 2000US-0190076P.  
 PR 18-APR-2000; 2000US-0198123P.  
 PR 19-MAY-2000; 2000US-0205515P.  
 PR 07-JUN-2000; 2000US-0209467P.  
 PR 28-JUN-2000; 2000US-0214886P.  
 PR 30-JUN-2000; 2000US-0215135P.  
 PR 07-JUL-2000; 2000US-0216647P.  
 PR 07-JUL-2000; 2000US-0216880P.  
 PR 11-JUL-2000; 2000US-0217487P.  
 PR 11-JUL-2000; 2000US-0217496P.  
 PR 14-JUL-2000; 2000US-0218290P.  
 PR 26-JUL-2000; 2000US-0220963P.  
 PR 26-JUL-2000; 2000US-0220964P.  
 PR 14-AUG-2000; 2000US-0224518P.  
 PR 14-AUG-2000; 2000US-0224519P.

PR 14-AUG-2000; 2000US-0225213P.  
 PR 14-AUG-2000; 2000US-0225214P.  
 PR 14-AUG-2000; 2000US-0225266P.  
 PR 14-AUG-2000; 2000US-0225267P.  
 PR 14-AUG-2000; 2000US-0225268P.  
 PR 14-AUG-2000; 2000US-0225270P.  
 PR 14-AUG-2000; 2000US-0225447P.  
 PR 14-AUG-2000; 2000US-0225757P.  
 PR 14-AUG-2000; 2000US-0225758P.  
 PR 14-AUG-2000; 2000US-0225759P.  
 PR 18-AUG-2000; 2000US-0226279P.  
 PR 22-AUG-2000; 2000US-0226681P.  
 PR 22-AUG-2000; 2000US-0226688P.  
 PR 23-AUG-2000; 2000US-0227009P.  
 PR 30-AUG-2000; 2000US-0228924P.  
 PR 01-SEP-2000; 2000US-0229287P.  
 PR 01-SEP-2000; 2000US-0229343P.  
 PR 01-SEP-2000; 2000US-0229344P.  
 PR 01-SEP-2000; 2000US-0229345P.  
 PR 05-SEP-2000; 2000US-0229509P.  
 PR 05-SEP-2000; 2000US-0229513P.  
 PR 06-SEP-2000; 2000US-0230437P.  
 PR 06-SEP-2000; 2000US-0230438P.  
 PR 08-SEP-2000; 2000US-0231242P.  
 PR 08-SEP-2000; 2000US-0231243P.  
 PR 08-SEP-2000; 2000US-0231244P.  
 PR 08-SEP-2000; 2000US-0231413P.  
 PR 08-SEP-2000; 2000US-0231414P.  
 PR 08-SEP-2000; 2000US-0232080P.  
 PR 08-SEP-2000; 2000US-0232081P.  
 PR 12-SEP-2000; 2000US-0231968P.  
 PR 14-SEP-2000; 2000US-0232397P.  
 PR 14-SEP-2000; 2000US-0232398P.  
 PR 14-SEP-2000; 2000US-0232399P.  
 PR 14-SEP-2000; 2000US-0232400P.  
 PR 14-SEP-2000; 2000US-0232401P.  
 PR 14-SEP-2000; 2000US-0233063P.  
 PR 14-SEP-2000; 2000US-0233064P.  
 PR 14-SEP-2000; 2000US-0233065P.  
 PR 21-SEP-2000; 2000US-0234223P.  
 PR 21-SEP-2000; 2000US-0234224P.  
 PR 25-SEP-2000; 2000US-0234997P.  
 PR 25-SEP-2000; 2000US-0234998P.  
 PR 26-SEP-2000; 2000US-0235484P.  
 PR 27-SEP-2000; 2000US-0235834P.  
 PR 27-SEP-2000; 2000US-0235836P.  
 PR 29-SEP-2000; 2000US-0236327P.  
 PR 29-SEP-2000; 2000US-0236328P.  
 PR 29-SEP-2000; 2000US-0236367P.  
 PR 29-SEP-2000; 2000US-0236368P.  
 PR 29-SEP-2000; 2000US-0236369P.  
 PR 29-SEP-2000; 2000US-0236370P.  
 PR 02-OCT-2000; 2000US-0236802P.  
 PR 02-OCT-2000; 2000US-0237037P.  
 PR 02-OCT-2000; 2000US-0237038P.  
 PR 02-OCT-2000; 2000US-0237039P.  
 PR 13-OCT-2000; 2000US-0237040P.  
 PR 13-OCT-2000; 2000US-0239935P.  
 PR 13-OCT-2000; 2000US-0239937P.  
 PR 20-OCT-2000; 2000US-0240960P.  
 PR 20-OCT-2000; 2000US-0241221P.  
 PR 20-OCT-2000; 2000US-0241785P.  
 PR 20-OCT-2000; 2000US-0241786P.  
 PR 20-OCT-2000; 2000US-0241787P.  
 PR 20-OCT-2000; 2000US-0241808P.  
 PR 20-OCT-2000; 2000US-0241809P.  
 PR 20-OCT-2000; 2000US-0241826P.  
 PR 01-NOV-2000; 2000US-0244617P.  
 PR 08-NOV-2000; 2000US-0246474P.  
 PR 08-NOV-2000; 2000US-0246475P.  
 PR 08-NOV-2000; 2000US-0246476P.  
 PR 08-NOV-2000; 2000US-0246477P.  
 PR 08-NOV-2000; 2000US-0246478P.



```

PR 05-SEP-2000; 2000US-0229513P.
PR 06-SEP-2000; 2000US-0230437P.
PR 06-SEP-2000; 2000US-0230438P.
PR 08-SEP-2000; 2000US-0231242P.
PR 08-SEP-2000; 2000US-0231243P.
PR 08-SEP-2000; 2000US-0231244P.
PR 08-SEP-2000; 2000US-0231413P.
PR 08-SEP-2000; 2000US-0231414P.
PR 08-SEP-2000; 2000US-0232080P.
PR 12-SEP-2000; 2000US-0232081P.
PR 14-SEP-2000; 2000US-0232377P.
PR 14-SEP-2000; 2000US-0232398P.
PR 14-SEP-2000; 2000US-0232399P.
PR 14-SEP-2000; 2000US-0232400P.
PR 14-SEP-2000; 2000US-0232401P.
PR 14-SEP-2000; 2000US-0233063P.
PR 14-SEP-2000; 2000US-0233064P.
PR 14-SEP-2000; 2000US-0233065P.
PR 21-SEP-2000; 2000US-0234223P.
PR 21-SEP-2000; 2000US-0234274P.
PR 25-SEP-2000; 2000US-0234997P.
PR 25-SEP-2000; 2000US-0234998P.
PR 26-SEP-2000; 2000US-0235484P.
PR 27-SEP-2000; 2000US-0235834P.
PR 27-SEP-2000; 2000US-0235836P.
PR 29-SEP-2000; 2000US-0236327P.
PR 29-SEP-2000; 2000US-0236367P.
PR 29-SEP-2000; 2000US-0236368P.
PR 29-SEP-2000; 2000US-0236369P.
PR 29-SEP-2000; 2000US-0236370P.
PR 29-SEP-2000; 2000US-0236802P.
PR 02-OCT-2000; 2000US-0237037P.
PR 02-OCT-2000; 2000US-0237038P.
PR 02-OCT-2000; 2000US-0237039P.
PR 02-OCT-2000; 2000US-0237040P.
PR 13-OCT-2000; 2000US-0239335P.
PR 13-OCT-2000; 2000US-0239337P.
PR 20-OCT-2000; 2000US-0240960P.
PR 20-OCT-2000; 2000US-0241221P.
PR 20-OCT-2000; 2000US-0241785P.
PR 20-OCT-2000; 2000US-0241786P.
PR 20-OCT-2000; 2000US-0241787P.
PR 20-OCT-2000; 2000US-0241808P.
PR 20-OCT-2000; 2000US-0241809P.
PR 01-NOV-2000; 2000US-0244817P.
PR 08-NOV-2000; 2000US-0246474P.
PR 08-NOV-2000; 2000US-0246475P.
PR 08-NOV-2000; 2000US-0246476P.
PR 08-NOV-2000; 2000US-0246477P.
PR 08-NOV-2000; 2000US-0246478P.
PR 08-NOV-2000; 2000US-0246523P.
PR 08-NOV-2000; 2000US-0246524P.
PR 08-NOV-2000; 2000US-0246525P.
PR 08-NOV-2000; 2000US-0246526P.
PR 08-NOV-2000; 2000US-0246527P.
PR 08-NOV-2000; 2000US-0246528P.
PR 08-NOV-2000; 2000US-0246532P.
PR 08-NOV-2000; 2000US-0246609P.
PR 08-NOV-2000; 2000US-0246610P.
PR 08-NOV-2000; 2000US-0246611P.
PR 08-NOV-2000; 2000US-0246612P.
PR 17-NOV-2000; 2000US-0249207P.
PR 17-NOV-2000; 2000US-0249208P.
PR 17-NOV-2000; 2000US-0249209P.
PR 17-NOV-2000; 2000US-0249210P.
PR 17-NOV-2000; 2000US-0249211P.
PR 17-NOV-2000; 2000US-0249212P.
PR 17-NOV-2000; 2000US-0249213P.
PR 17-NOV-2000; 2000US-0249214P.
PR 17-NOV-2000; 2000US-0249215P.
PR 17-NOV-2000; 2000US-0249216P.

PR 17-NOV-2000; 2000US-0249217P.
PR 17-NOV-2000; 2000US-0249218P.
PR 17-NOV-2000; 2000US-0249244P.
PR 17-NOV-2000; 2000US-0249245P.
PR 17-NOV-2000; 2000US-0249264P.
PR 17-NOV-2000; 2000US-0249265P.
PR 17-NOV-2000; 2000US-0249297P.
PR 17-NOV-2000; 2000US-0249299P.
PR 01-DEC-2000; 2000US-0249300P.
PR 01-DEC-2000; 2000US-0250160P.
PR 05-DEC-2000; 2000US-0250391P.
PR 05-DEC-2000; 2000US-0251030P.
PR 05-DEC-2000; 2000US-b251988P.
PR 06-DEC-2000; 2000US-0256719P.
PR 06-DEC-2000; 2000US-0251479P.
PR 08-DEC-2000; 2000US-0251856P.
PR 08-DEC-2000; 2000US-0251868P.
PR 08-DEC-2000; 2000US-0251869P.
PR 08-DEC-2000; 2000US-0251989P.
PR 11-DEC-2000; 2000US-0251990P.
PR 05-JAN-2001; 2001US-0254097P.
PR 05-JAN-2001; 2001US-0259678P.
XX
XX
PA (HUMA-) HUMAN GENOME SCI INC.
PI Rosen CA, Barash SC, Ruben SM;
XX
XX WPI; 2001-483426/52.
DR P-PSDB; AAM83591.
XX
XX Nucleic acids encoding human immune/hematopoietic antigen polypeptides,
PT useful for preventing, diagnosing and/or treating cancers and metastasis.
XX
XX Claim 1; SEQ ID NO 1432; 3071pp + Sequence Listing; English.
XX
XX AAK54951 to AAK64702 encode the human immune/hematopoietic antigen (I)
CC amino acid sequences given in AAM82170 to AAM91921. (I) have cytostatic
CC activity, and can be used in gene therapy and vaccine production. (I)
CC proteins and polynucleotides may be used in the prevention, diagnosis and
CC treatment of diseases associated with inappropriate (I) expression. For
CC example, they may be used to treat disorders associated with decreased
CC expression by rectifying mutations or deletions in a patient's genome
CC that affect the activity of (I) by expressing inactive proteins or to
CC supplement the patients own production of (I). Additionally, (I)
CC polynucleotides may be used to produce the secreted (I), by inserting the
CC nucleic acids into a host cell and culturing the cell to express the
CC protein. (I) proteins and polynucleotides may be used to prevent,
CC diagnose and treat immune/hematopoietic-related diseases, especially
CC cancers and cancer metastases of hematopoietic-derived cells. AAK64703
CC to AAK87694 represent human immune/hematopoietic antigen genomic
CC sequences from the present invention. AAK54942 to AAK54950 and AAM82169
CC represent sequences used in the exemplification of the present invention
XX
SQ Sequence 359 BP; 122 A; 78 C; 73 G; 84 T; 0 U; 2 Other;

Query Match 78.2%; Score 17.2; DB 4; Length 359;
Best Local Similarity 86.4%; Pred. No. 2.1e+02;
Matches 19; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

Qy 1 TGAGAGAGTTCTGGGTGTCCTA 22
Db 148 TGAGTGAGTTCTGGCTCTCCTA 127

Search completed: February 19, 2004, 21:51:47
Job time : 24.5821 secs

```

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM nucleic - nucleic search, using sw model  
Run on: February 19, 2004, 21:21:54 ; Search time 145.664 Seconds  
(without alignments)  
4510.152 Million cell updates/sec

Title: US-09-939-853A-142  
Perfect score: 22  
Sequence: 1 tggagagattctgggtgccta 22

Scoring table: IDENTITY\_NUC  
Gapop 10.0 , Gapext 1.0

Searched: 27513289 seqs, 14931090276 residues

Total number of hits satisfying chosen parameters: 55026578

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database :

EST:\*

1: em\_estba:\*

2: em\_esthum:\*

3: em\_estin:\*

4: em\_estmu:\*

5: em\_estov:\*

6: em\_estpl:\*

7: em\_estro:\*

8: em\_hic:\*

9: gb\_estl:\*

10: gb\_est2:\*

11: gb\_hic:\*

12: gb\_est3:\*

13: gb\_est4:\*

14: gb\_est5:\*

15: em\_estfun:\*

16: em\_estom:\*

17: em\_gss\_hum:\*

18: em\_gss\_inv:\*

19: em\_gss\_pln:\*

20: em\_gss\_vrt:\*

21: em\_gss\_fun:\*

22: em\_gss\_mam:\*

23: em\_gss\_mus:\*

24: em\_gss\_pro:\*

25: em\_gss\_rtd:\*

26: em\_gss\_phg:\*

27: em\_gss\_vrl:\*

28: gb\_gsl:\*

29: gb\_gsl2:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Match	Length	ID	Description
1	22	100.0	616	13	BX383606
2	22	100.0	778	12	BG178487
3	22	100.0	878	12	BQ053486
4	22	100.0	986	12	BQ054265

5	22	100.0	1020	12	BQ054281
6	22	100.0	1201	9	AL541041
7	18.8	85.5	573	28	CC322697
8	18.4	83.6	598	28	AZ827349
9	18.4	83.6	651	9	AI727541
10	18.4	83.6	2919	11	AK035058
11	17.8	80.9	283	28	AZ940479
12	17.8	80.9	296	29	CC513479
13	17.8	80.9	464	28	BH834762
14	17.8	80.9	498	14	CF606764
15	17.8	80.9	601	28	BH826297
16	17.8	80.9	652	12	BJ011310
17	17.8	80.9	690	9	AU168324
18	17.8	80.9	699	9	AV846834
19	17.8	80.9	780	28	AZ727518
20	17.8	80.9	910	14	CA965326
21	17.8	80.9	1062	28	CC305491
22	17.8	80.9	1066	28	CC250043
23	17.4	79.1	290	13	BX500783
24	17.4	79.1	295	12	BI430374
25	17.4	79.1	438	9	AA128423
26	17.4	79.1	537	9	AV598294
27	17.4	79.1	558	28	AZ749071
28	17.4	79.1	583	29	CE551267
29	17.4	79.1	583	13	BU065505
30	17.4	79.1	588	13	BU064867
31	17.4	79.1	925	14	CF551819
32	17.4	79.1	1062	29	CNS0602K
33	17.4	79.1	1160	29	CNS06718
34	17.4	79.1	1878	29	AV419877
35	17.2	78.2	315	9	AA640882
36	17.2	78.2	363	10	BB730751
37	17.2	78.2	383	13	BY675867
38	17.2	78.2	389	29	CG578968
39	17.2	78.2	407	9	AV797883
40	17.2	78.2	408	10	BE654580
41	17.2	78.2	430	10	BE848016
42	17.2	78.2	431	9	AI149606
43	17.2	78.2	439	9	AA693188
44	17.2	78.2	470	9	AI616154
45	17.2	78.2	480	28	AZ2644203

## ALIGNMENTS

RESULT 1  
BX383606  
LOCUS  
DEFINITION BX383606 Homo sapiens T CELLS (JURKAT CELL LINE) COT 10-NORMALIZED  
Homo sapiens cDNA clone CS0DJ013YK10 5-PRIME, mRNA sequence.  
ACCESSION BX383606  
VERSION BX383606.1 GI:30457152  
KEYWORDS EST.  
SOURCE Homo sapiens (human)  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
REFERENCE 1 (bases 1 to 616)  
Li, W.B., Gruber, C., Jessee, J. and Polayes, D.  
Full-length cDNA libraries and normalization  
Unpublished (2001)  
JOURNAL Contact: Genoscope  
COMMENT Genoscope - Centre National de Sequencage  
BP 191 91006 EVRY cedex - France  
Email: seqref@genoscope.cns.fr, Web : www.genoscope.cns.fr  
Library was constructed by Life Technologies, a division of  
Invitrogen. This sequence belongs to sequence cluster 9825.r For  
more information about this cluster, see  
http://www.genoscope.cns.fr/  
cgi-bin/cluster.cgi?seq=CS0DJ013BF05QP1&cluster=9825.r. Contact :  
Feng Liang Email : fliang@lifetech.com URL :  
http://fulllength.invitrogen.com/ invitrogen Corporation 1600

Faraday Avenue Genoscope sequence ID : CS0DJ013BF05QPi.

## FEATURES

source

```
1. .616
Location/Qualifiers
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="CS0DJ013YK10"
/cell_type="T CELLS (JURKAT CELL LINE) COT 10-NORMALIZED"
/cell_line="JURKAT"
/clone_lib="Homo sapiens T CELLS (JURKAT CELL LINE) COT
10-NORMALIZED"
/note="1st strand cDNA was primed with a NotI-oligo(dT)
primer. Five prime end enriched, double-strand cDNA was
digested with Not I and cloned into the Not I and EcoR V
sites of the pCMVSPORT6 vector. Library was normalized."
```

## ORIGIN

```
Query Match 100.0%; Score 22; DB 13; Length 616;
Best Local Similarity 100.0%; Pred. No. 29;
Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

```
QY 1 TGAGAGAGTTCTGGGTGTCCTA 22
Db 273 TGAGAGAGTTCTGGGTGTCCTA 294
```

## RESULT 2

```
BGI78487
LOCUS 778 bp mRNA linear EST 06-FEB-2001
DEFINITION 60328305F1 NIH_MGC_91 Homo sapiens cDNA clone IMAGE:4429896 5',
mRNA sequence.
```

```
ACCESSION BGI78487
VERSION BGI78487.1 GI:12685190
```

```
KEYWORDS EST.
```

```
SOURCE Homo sapiens (human)
```

## ORGANISM

```
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
```

## REFERENCE

```
1 (bases 1 to 778)
NIH-MGC http://mgc.nci.nih.gov/.
National Institutes of Health, Mammalian Gene Collection (MGC)
Unpublished (1999).
```

## JOURNAL

```
COMMENT Contact: Robert Strausberg, Ph.D.
```

```
Tissue Procurement: DCTD/DPF
cDNA Library Preparation: Life Technologies, Inc.
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Incyte Genomics, Inc.
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
http://image.llnl.gov
```

```
Plate: LLMW0182 row: i column: 01
High quality sequence stop: 657.
```

## FEATURES

source

```
1. .778
Location/Qualifiers
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="IMAGE:4429896"
/tissue_type="adenocarcinoma, cell line"
/lab_host="DH10B (phage-resistant)"
/clone_lib="NIH MGC 91"
/note="Organ: prostate; Vector: pCMV-SPORT6; Site 1: NotI;
Site 2: SalI; Cloned unidirectionally; oligo-dT primed.
Average insert size 1.4 kb. Library enriched for
full-length clones and constructed by Life Technologies.
Note: this is a NIH_MGC Library."
```

## ORIGIN

```
Query Match 100.0%; Score 22; DB 12; Length 778;
Best Local Similarity 100.0%; Pred. No. 31;
Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

```
QY 1 TGAGAGAGTTCTGGGTGTCCTA 22
Db 188 TGAGAGAGTTCTGGGTGTCCTA 209
```

## RESULT 3

```
BQ053486
LOCUS 878 bp mRNA linear EST 29-MAR-2002
DEFINITION AGENCOURT_6822017 NIH_MGC_106 Homo sapiens cDNA clone IMAGE:5935253
5', mRNA sequence.
```

```
ACCESSION BQ053486
VERSION BQ053486.1 GI:19812826
```

## KEYWORDS

```
EST.
```

```
SOURCE Homo sapiens (human)
```

## ORGANISM

```
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
```

## REFERENCE

```
1 (bases 1 to 878)
NIH-MGC http://mgc.nci.nih.gov/.
```

## AUTHORS

```
National Institutes of Health, Mammalian Gene Collection (MGC)
Unpublished (1999)
```

## JOURNAL

```
COMMENT Contact: Robert Strausberg, Ph.D.
```

```
Email: cgabbs@mail.nih.gov
```

```
Tissue Procurement: Dr. Daniel McVicar, DBS/NCI
```

```
cDNA Library Preparation: Rubin Laboratory
```

```
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
```

```
DNA Sequencing by: Agencourt Bioscience Corporation
```

```
Clone distribution: MGC clone distribution information can be
```

```
found through the I.M.A.G.E. Consortium/LLNL at:
```

```
http://image.llnl.gov
```

```
Plate: LLM2122 row: 1 column: 06
```

```
High quality sequence stop: 394.
```

## FEATURES

source

```
1. 878
Location/Qualifiers
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="IMAGE:5935253"
/tissue_type="natural killer cells, cell line"
/lab_host="DH10B (phage-resistant)"
/clone_lib="NIH MGC 106"
/note="Organ: blood; Vector: pOTB7; Site 1: XhoI; Site 2:
EcoRI; cDNA made by oligo-dT priming. Directionally cloned
into EcoRI/XhoI sites using the following 5' adaptor:
GGCAGCAG(G). Library constructed by Ling Hong in the
Laboratory of Gerald M. Rubin (University of California,
Berkeley) using ZAP-cDNA synthesis kit (Stratagene) and
Superscript II RT (Life Technologies). Note: this is a
NIH_MGC Library."
```

## ORIGIN

```
Query Match 100.0%; Score 22; DB 12; Length 878;
Best Local Similarity 100.0%; Pred. No. 33;
Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

```
QY 1 TGAGAGAGTTCTGGGTGTCCTA 22
Db 99 TGAGAGAGTTCTGGGTGTCCTA 120
```

## RESULT 4

```
BQ054265
LOCUS 986 bp mRNA linear EST 29-MAR-2002
DEFINITION AGENCOURT_6830248 NIH_MGC_106 Homo sapiens cDNA clone IMAGE:5935339
5', mRNA sequence.
```

```
ACCESSION BQ054265
VERSION BQ054265.1 GI:19813605
```

## KEYWORDS

```
EST.
```

```
SOURCE Homo sapiens (human)
```

## ORGANISM

```
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
```

## REFERENCE

```
1 (bases 1 to 986)
```

AUTHORS NIH-MGC <http://mgc.nci.nih.gov/>  
 TITLE National Institutes of Health, Mammalian Gene Collection (MGC)  
 JOURNAL Unpublished (1999)  
 COMMENT Contact: Robert Strausberg, Ph.D.  
 Email: [cgapsb-remail.nih.gov](mailto:cgapsb-remail.nih.gov)  
 Tissue Procurement: Dr. Daniel McVicar, DBS/NCI  
 cDNA Library Preparation: Rubin Laboratory  
 cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)  
 DNA Sequencing by: Agencourt Bioscience Corporation  
 Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at:  
<http://image.llnl.gov>  
 Plate: L1CM2125 row: i column: 12  
 High quality sequence stop: 515.

## FEATURES

source  
 1. .986  
 /organism="Homo sapiens"  
 /mol\_type="mRNA"  
 /db\_xref="taxon:9606"  
 /clone="IMAGE:5936339"  
 /tissue\_type="natural killer cells, cell line"  
 /lab\_host="DH10B (phage-resistant)"  
 /clone\_lib="NIH\_MGC\_106"  
 /note="Organ: Blood; Vector: pOTB7; Site: 1: XhoI; Site 2: EcoRI; cDNA made by oligo-dT priming. Directionally cloned into EcoRI/XhoI sites using the following 5' adaptor: GGCACGAG(G). Library constructed by Ling Hong in the laboratory of Gerald M. Rubin (University of California, Berkeley) using ZAP-cDNA synthesis kit (Stratagene) and Superscript II RT (Life Technologies). Note: this is a NIH\_MGC Library."

## ORIGIN

Query Match 100.0%; Score 22; DB 12; Length 986;  
 Best Local Similarity 100.0%; Pred. No. 34;  
 Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 TGAGAGAGTTCTGGGTGTCCTA 22  
 |||||  
 Db 105 TGAGAGAGTTCTGGGTGTCCTA 126

RESULT 5  
 BQ054281  
 LOCUS AGENCOURT\_6830234 NIH\_MGC\_106 Homo sapiens cDNA clone IMAGE:5936362  
 DEFINITION 5', mRNA sequence.  
 ACCESSION BQ054281  
 VERSION BQ054281.1 GI:19813621  
 KEYWORDS EST  
 SOURCE Homo sapiens (human)  
 ORGANISM  
 Eukaryota; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
 REFERENCE 1 (bases 1 to 1020)  
 AUTHORS NIH-MGC <http://mgc.nci.nih.gov/>  
 TITLE National Institutes of Health, Mammalian Gene Collection (MGC)  
 JOURNAL Unpublished (1999)  
 COMMENT Contact: Robert Strausberg, Ph.D.  
 Email: [cgapsb-remail.nih.gov](mailto:cgapsb-remail.nih.gov)  
 Tissue Procurement: Dr. Daniel McVicar, DBS/NCI  
 cDNA Library Preparation: Rubin Laboratory  
 cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)  
 DNA Sequencing by: Agencourt Bioscience Corporation  
 Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at:  
<http://image.llnl.gov>  
 Plate: L1CM2125 row: j column: 11  
 High quality sequence stop: 556.

## FEATURES

source  
 1. .1020  
 /organism="Homo sapiens"  
 /mol\_type="mRNA"

/db\_xref="taxon:9606"  
 /clone="IMAGE:5936362"  
 /tissue\_type="natural killer cells, cell line"  
 /lab\_host="DH10B (phage-resistant)"  
 /clone\_lib="NIH\_MGC\_106"  
 /note="Organ: blood; Vector: pOTB7; Site 1: XhoI; Site 2: EcoRI; cDNA made by oligo-dT priming. Directionally cloned into EcoRI/XhoI sites using the following 5' adaptor: GGCACGAG(G). Library constructed by Ling Hong in the laboratory of Gerald M. Rubin (University of California, Berkeley) using ZAP-cDNA synthesis kit (Stratagene) and Superscript II RT (Life Technologies). Note: this is a NIH\_MGC Library."

## ORIGIN

Query Match 100.0%; Score 22; DB 12; Length 1020;  
 Best Local Similarity 100.0%; Pred. No. 34;  
 Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 TGAGAGAGTTCTGGGTGTCCTA 22  
 |||||  
 Db 38 TGAGAGAGTTCTGGGTGTCCTA 59

RESULT 6  
 AL541041  
 LOCUS AL541041 Homo sapiens PLACENTA Homo sapiens cDNA clone CS0DE005YK23  
 DEFINITION 5-PRIME, mRNA sequence.  
 ACCESSION AL541041  
 VERSION AL541041.2 GI:30544829  
 KEYWORDS EST.

## ORIGIN

source  
 1. .1201  
 /organism="Homo sapiens"  
 /mol\_type="mRNA"  
 /db\_xref="taxon:9606"  
 /clone="CS0DE005YK23"  
 /tissue\_type="PLACENTA"  
 /note="Vector: pCMVSPORT 6; 1st strand cDNA was primed with a NotI-oligo(dT) primer. Five prime end enriched, double-strand cDNA was digested with Not I and cloned into the Not I and EcoRV sites of the pCMVSPORT 6 vector. Library was not normalized."  
 ORGANISM  
 Homo sapiens (human)  
 Eukaryota; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
 REFERENCE 1 (bases 1 to 1201)  
 AUTHORS Li, W.B., Gruber, C., Jessee, J. and Polayes, D.  
 TITLE Full-length cDNA libraries and normalization  
 JOURNAL Unpublished (2001)  
 COMMENT On Feb 15, 2001 this sequence version replaced gi:12871733.  
 CONTACT: Genoscope  
 Genoscope - Centre National de Sequencage  
 BP 191 91006 EVRY cedex - France  
 Email: [seqreg@genoscope.cns.fr](mailto:seqreg@genoscope.cns.fr), Web : [www.genoscope.cns.fr](http://www.genoscope.cns.fr)  
 Library was constructed by Life Technologies, a division of Invitrogen. This sequence belongs to sequence cluster 9825.r For more information about this cluster, see <http://www.genoscope.cns.fr/cgi-bin/cluster.cgi?seq=CS0DE005AF12QPI&cluster=9825.r>. Contact : Feng Liang Email : [fliang@lifetech.com](mailto:fliang@lifetech.com) URL : Corporation 1600 Faraday Avenue Genoscope sequence ID : CS0DE005AF12QPI.  
 Location/Qualifiers  
 1. .1201

## FEATURES

source  
 1. .1201  
 /organism="Homo sapiens"  
 /mol\_type="mRNA"  
 /db\_xref="taxon:9606"  
 /clone="CS0DE005YK23"  
 /tissue\_type="PLACENTA"  
 /note="Vector: pCMVSPORT 6; 1st strand cDNA was primed with a NotI-oligo(dT) primer. Five prime end enriched, double-strand cDNA was digested with Not I and cloned into the Not I and EcoRV sites of the pCMVSPORT 6 vector. Library was not normalized."  
 ORGANISM  
 Homo sapiens (human)  
 Eukaryota; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
 REFERENCE 1 (bases 1 to 1201)  
 AUTHORS Li, W.B., Gruber, C., Jessee, J. and Polayes, D.  
 TITLE Full-length cDNA libraries and normalization  
 JOURNAL Unpublished (2001)  
 COMMENT On Feb 15, 2001 this sequence version replaced gi:12871733.  
 CONTACT: Genoscope  
 Genoscope - Centre National de Sequencage  
 BP 191 91006 EVRY cedex - France  
 Email: [seqreg@genoscope.cns.fr](mailto:seqreg@genoscope.cns.fr), Web : [www.genoscope.cns.fr](http://www.genoscope.cns.fr)  
 Library was constructed by Life Technologies, a division of Invitrogen. This sequence belongs to sequence cluster 9825.r For more information about this cluster, see <http://www.genoscope.cns.fr/cgi-bin/cluster.cgi?seq=CS0DE005AF12QPI&cluster=9825.r>. Contact : Feng Liang Email : [fliang@lifetech.com](mailto:fliang@lifetech.com) URL : Corporation 1600 Faraday Avenue Genoscope sequence ID : CS0DE005AF12QPI.  
 Location/Qualifiers  
 1. .1201

## ORIGIN

Query Match 100.0%; Score 22; DB 9; Length 1201;  
 Best Local Similarity 100.0%; Pred. No. 36;  
 Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 TGAGAGAGTTCTGGGTGTCCTA 22

```

Db      285 TGAGAGAGTTCTGGGTGTCCTA 306
|||||
RESULT 7
CC322697
LOCUS    573 bp      DNA      linear      GSS 14-MAY-2003
DEFINITION  TAM32-33G24_Sp6.1 TAM32 Gallus gallus genomic clone TAM32-33G24,
genomic survey sequence.
ACCESSION  CC322697
VERSION    CC322697.1 GI:30716755
KEYWORDS   GSS.
SOURCE     Gallus gallus (chicken)
ORGANISM   Gallus gallus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Archosauria; Aves; Neognathae; Galliformes; Phasianidae;
Phasianinae; Gallus.
1 (bases 1 to 573)
Kremitzki, C., Higginbotham, J., Wylie, K., Carter, J., McPherson, J.,
Warren, W., Graves, T., Mardis, E. and Wilson, R.
Gallus gallus BAC End Reads
Unpublished (2003)
Contact: Richard K. Wilson
Genome Sequencing Center
Washington University School of Medicine
Email: submissions@watson.wustl.edu
Insert Length: 152000 Std Error: 0.00
Seq primer: Sp6 ATTAGGTGACATATAG
Class: BAC ends
High quality sequence start: 3
High quality sequence stop: 478.
Location/Qualifiers
1..573
/organism="Gallus gallus"
/mol_type="genomic DNA"
/strain="Red Jungle Fowl"
/db_xref="taxon:9031"
/clone="TAM32-33G24"
/sex="female"
/cell_line="UCD001, inbred 256"
/clone_lib="TAM32"
/notes="Vector: pSCBAC1; Site 1: EcoRI; Site 2: EcoRI;
TAM32 Female Chicken library_ for library and clone
ordering information: http://www.hbz.tamu.edu"

ORIGIN
Query Match      85.5%; Score 18.8; DB 28; Length 573;
Best Local Similarity 90.9%; Pred. No. 7.5e+02;
Matches 20; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 TGAGAGAGTTCTGGGTGTCCTA 22
|||||
DB 361 TGAGAGAGTTCTGGGTGTCCTA 382
|||||

RESULT 8
AZ827349
LOCUS    598 bp      DNA      linear      GSS 20-FEB-2001
DEFINITION  ZM0103B18R Mouse 10kb plasmid UUC1M library Mus musculus genomic
clone UUC2M0103B18 R, genomic survey sequence.
ACCESSION  AZ827349
VERSION    AZ827349.1 GI:12997257
KEYWORDS   GSS.
SOURCE     Mus musculus (house mouse)
ORGANISM   Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
1 (bases 1 to 598)
Dunn, D., Aoyagi, A., Barber, M., Beacorn, T., Duval, B., Hamil, C.,
Islam, H., Longacre, S., Mahmood, M., Meenen, E., Pedersen, T.,
Reilly, M., Rose, M., Rose, R., Stokes, R., Tingey, A., von
Niederhausern, A. and Wright, D., Weiss, R.
Mouse whole genome scaffolding with paired end reads from 10kb

TITLE

```

```

plasmid inserts
Unpublished (2000)
Contact: Robert B. Weiss
University of Utah
Genome Center
Rm. 308, Biomedical Polymers Research Bldg., 20 S. 2030 E., SLC, UT
84112, USA
Tel: 801 585 5606
Fax: 801 585 7177
Email: dunn@genetics.utah.edu
Insert Length: 10000 Std Error: 0.00
Plate: 0103 row: B column: 18
Seq primer: CACACAGGAAACAGCTATGACC
Class: plasmid ends
High quality sequence stop: 598.
Location/Qualifiers
1..598
/organism="Mus musculus"
/mol_type="genomic DNA"
/strain="C57BL/6J"
/db_xref="taxon:10090"
/clone="UUC2M0103B18"
/sex="Male"
/lab_host="E. Coli strain XL10-Gold, Tl-resistant, F-"
/clone_lib="Mouse 10kb plasmid UUC1M library"
/notes="Vector: PWD42nv; Purified genomic DNA from M.
musculus C57BL/6J (male) was obtained from the Jackson
Laboratory Mouse DNA Resource
(http://www.jax.org/resources/documents/dnares/). The DNA
was hydrodynamically sheared by repeated passage through a
0.005 inch orifice at constant velocity. The sheared DNA
was blunt end-repaired with T4 DNA polymerase and T4
polynucleotide kinase. Adaptor oligonucleotides were
ligated to the blunt ends in high molar excess. The
adaptor DNA was purified and size-selected for a 9.5 to
10.5 kb range using preparative agarose gel
electrophoresis. Vector DNA was prepared from a derivative
of PWD42 (gi|4732114|gb|AF129072.1), a copy-number
inducible derivative of plasmid R1. The vector was ligated
with adaptors complementary to the insert adaptors and
purified. The sheared, adaptor mouse DNA was annealed to
adaptor vector DNA, and transformed into
chemically-competent E. coli XL10-Gold (Stratagene) cells
and selected for ampicillin resistance."

ORIGIN
Query Match      83.6%; Score 18.4; DB 28; Length 598;
Best Local Similarity 95.0%; Pred. No. 1.1e+03;
Matches 19; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 TGAGAGAGTTCTGGGTGTC 20
|||||
DB 279 TGAGAGAGTTCTGGGTGTC 298
|||||

RESULT 9
AI727541
LOCUS    651 bp      mRNA      linear      EST 11-JUN-1999
DEFINITION  BNLGH1847 Six-day Corton fiber Gossypium hirsutum cDNA 5' similar
to (AF039373) polyprotein [Arabidopsis thaliana], mRNA sequence.
ACCESSION  AI727541
VERSION    AI727541.1 GI:5046393
KEYWORDS   EST.
SOURCE     Gossypium hirsutum (upland cotton)
ORGANISM   Gossypium hirsutum
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
Rosids; eurosids II; Malvales; Malvaceae; Malvoideae; Gossypium.
1 (bases 1 to 651)
Blewitt, M., Matz, B.C., Davy, D.F. and Burr, B.
ESTs from developing cotton fiber
Unpublished (1999)
Contact: Ben Burr

```



Biology Department  
Brookhaven National Laboratory  
Upton, NY 11973, USA  
Tel: 516-344-3396  
Fax: 516-344-3407  
Email: burr@bnl.bnl.gov  
Seq primer: T3 Primer.  
Location/Qualifiers

# FEATURES

source  
1. .651  
/organism="Gossypium hirsutum"  
/mol\_type="mRNA"  
/cultivar="Acala Maxxa"  
/db\_xref="taxon:3635"  
/tissue\_type="Immature fiber"  
/dev\_stage="Six days post anthesis"  
/lab\_host="XLI-Blue"  
/clone\_lib="Six-day Cotton fiber"  
/notes="Vector: pBluescript II KS+"

## ORIGIN

Query Match 83.6%; Score 18.4; DB 9; Length 651;  
Best Local Similarity 95.0%; Pred. No. 1.2e+03;  
Matches 19; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 TGAGAGAGTCTCGGTCTCC 20

Db 91 TGAGAGAGTCTCGGTCTCC 110

## RESULT 10

AK035058

## LOCUS

DEFINITION

AK035058 2919 bp mRNA linear HTC 18-SEP-2003  
Mus musculus 12 days embryo embryonic body between diaphragm region  
and neck cDNA, RIKEN full-length enriched library, clone:9430080I20  
product:hypothetical protein, full insert sequence.

## ACCESSION

VERSION AK035058.1 GI:26084370

KEYWORDS HTC; CAP trapper.

## SOURCE

Mus musculus (house mouse)

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

## REFERENCE

1

Carninci, P. and Hayashizaki, Y.

High-efficiency full-length cDNA cloning

Meth. Enzymol. 303, 19-44 (1999)

99279253

MEDLINE

10349636

## REFERENCE

2

Carninci, P., Shibata, Y., Hayatsu, N., Sugahara, Y., Shibata, K.,  
Itoh, M., Konno, H., Okazaki, Y., Muramatsu, M. and Hayashizaki, Y.

Normalization and subtraction of cap-trapper-selected cDNAs to

prepare full-length cDNA libraries for rapid discovery of new genes

Genome Res. 10 (10), 1617-1630 (2000)

20493374

MEDLINE

11042159

## REFERENCE

3

Shibata, K., Itoh, M., Aizawa, K., Nagaoka, S., Sasaki, N., Carninci, P.,  
Konno, H., Akiyama, J., Nishi, K., Kitsumai, T., Tashiro, H., Itoh, M.,

Sumi, N., Ishii, Y., Nakamura, S., Hazama, M., Nishine, T., Harada, A.,

Yamamoto, R., Matsumoto, H., Sakaguchi, S., Ikegami, T., Kashiwagi, K.,

Fujiwaka, S., Inoue, K., Togawa, Y., Izawa, M., Ohara, E., Watahiki, M.,

Yoneda, Y., Ishikawa, T., Ozawa, K., Tanaka, T., Matsura, S., Kawai, J.,

Okazaki, Y., Muramatsu, M., Inoue, Y., Kira, A. and Hayashizaki, Y.

RIKEN integrated sequence analysis (RISA) system--384-format

sequencing pipeline with 384 multicapillary sequencer

Genome Res. 10 (11), 1757-1771 (2000)

20530913

MEDLINE

11076861

## REFERENCE

4

The RIKEN Genome Exploration Research Group Phase II Team and the

FANTOM Consortium.

Functional annotation of a full-length mouse cDNA collection

## JOURNAL

REFERENCE

AUTHORS

TITLE

JOURNAL

REFERENCE

AUTHORS

Location/Qualifiers

1. .651

/organism="Gossypium hirsutum"

/mol\_type="mRNA"

/cultivar="Acala Maxxa"

/db\_xref="taxon:3635"

/tissue\_type="Immature fiber"

/dev\_stage="Six days post anthesis"

/lab\_host="XLI-Blue"

/clone\_lib="Six-day Cotton fiber"

/notes="Vector: pBluescript II KS+"

## TITLE

JOURNAL

REFERENCE

AUTHORS

TITLE

JOURNAL

REFERENCE

AUTHORS

TITLE

JOURNAL

REFERENCE

AUTHORS

TITLE

JOURNAL

REFERENCE

AUTHORS

TITLE

JOURNAL

REFERENCE

AUTHORS

TITLE

JOURNAL

REFERENCE

AUTHORS

TITLE

JOURNAL

REFERENCE

AUTHORS

TITLE

JOURNAL

REFERENCE

AUTHORS

TITLE

JOURNAL

REFERENCE

AUTHORS

TITLE

JOURNAL

REFERENCE

AUTHORS

TITLE

JOURNAL

REFERENCE

AUTHORS

TITLE

JOURNAL

REFERENCE

AUTHORS

TITLE

JOURNAL

REFERENCE

AUTHORS

TITLE

JOURNAL

REFERENCE

AUTHORS

TITLE

JOURNAL

REFERENCE

AUTHORS

TITLE

Nature 409, 685-690 (2001)

5

The PANTOM Consortium and the RIKEN Genome Exploration Research  
Group Phase I & II Team.

Analysis of the mouse transcriptome based on functional annotation

of 60,770 full-length cDNAs

Nature 420, 563-573 (2002)

6 (bases 1 to 2919)

Adachi, J., Aizawa, K., Akimura, T., Arakawa, T., Bono, H., Carninci, P.,

Fukuda, S., Furuno, M., Hanagaki, T., Hara, A., Hashizume, W.,

Hayashida, K., Hayatsu, N., Hiramoto, K., Hiraoka, T., Hirozane, T.,

Hori, F., Imotani, K., Ishii, Y., Itoh, M., Kagawa, I., Kasukawa, T.,

Kato, H., Kawai, J., Kojima, Y., Kondo, S., Konno, H., Kouda, M.,

Koya, S., Kurihara, C., Matsuyama, T., Miyazaki, A., Murata, M.,

Nakamura, M., Nishi, K., Nomura, K., Numazaki, R., Ohno, M., Ohsato, N.,

Okazaki, Y., Saito, R., Saitoh, H., Sakai, C., Sakai, K., Sakazume, N.,

Sano, H., Sasaki, D., Shibata, K., Shinagawa, A., Shiraki, T.,

Sogabe, Y., Tagami, M., Tagawa, A., Takahashi, F., Takaku-Akahira, S.,

Takeda, Y., Tanaka, T., Tomaru, A., Toya, T., Yasunishi, A.,

Muramatsu, M. and Hayashizaki, Y.

Direct Submission

Submitted (16-JUL-2001) Yoshihide Hayashizaki, The Institute of

Physical and Chemical Research (RIKEN), Laboratory for Genome

Exploration Research Group, RIKEN Genomic Sciences Center (GSC),

RIKEN Yokohama Institute; 1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama,

Kanagawa 230-0045, Japan (E-mail: genome-res@gsc.riken.go.jp,

URL: http://genome.gsc.riken.go.jp/, Tel: 81-45-503-9222,

Fax: 81-45-503-9216)

cDNA library was prepared and sequenced in Mouse Genome

Encyclopedia Project of Genome Exploration Research Group in Riken

Genomic Sciences Center and Genome Science Laboratory in RIKEN.

Division of Experimental Animal Research in Riken contributed to

prepare mouse tissues.

Please visit our web site for further details.

URL: http://genome.gsc.riken.go.jp/

URL: http://fantom.gsc.riken.go.jp/

Location/Qualifiers

1. .2919

/organism="Mus musculus"

/mol\_type="mRNA"

/strain="C57BL/6J"

/db\_xref="FANTOM\_DB:9430080I20"

/db\_xref="MGI:2399272"

/db\_xref="taxon:10090"

/clone="9430080I20"

/tissue\_type="embryonic body between diaphragm region and

neck"

/clone\_lib="RIKEN full-length enriched mouse cDNA library"

/dev\_stage="12 days embryo"

misc\_feature 1. .2919

/note="hypothetical protein (evidence: ProCrest)"

ORIGIN

Query Match 83.6%; Score 18.4; DB 11; Length 2919;

Best Local Similarity 95.0%; Pred. No. 1.9e+03;

Matches 19; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 2 GAGAGAGTCTCGGTCTCC 21

Db 589 GAGAGAGTCTCGGTCTCC 608

RESULT 11

AZ940479

LOCUS 2M0199D23R Mouse 10kb plasmid UGCG2M library Mus musculus genomic

clone UGCG2M0199D23 R, genomic survey sequence.

DEFINITION

ACCESSION AZ940479

VERSION AZ940479.1 GI:13802090

KEYWORDS GSS.

SOURCE Mus musculus (house mouse)

ORGANISM Mus musculus

Eukaryota; Chordata; Craniata; Vertebrata; Euteleostomi;

283 bp DNA linear GSS 26-APR-2001

REFERENCE 1 (bases 1 to 283)  
 AUTHORS Durn,D., Aoyagi,A., Barber,M., Beacorn,T., Duval,B., Hamil,C., Islam,H., Longacre,S., Mahmoud,M., Meenen,E., Pedersen,T., Reilly,M., Rose,M., Rose,R., Stokes,R., Tingey,A., von Niederhausen,A. and Wright,D. Weiss,R.  
 TITLE Mouse whole genome scaffolding with paired end reads from 10kb plasmid inserts  
 JOURNAL Unpublished (2000)  
 COMMENT Contact: Robert B. Weiss  
 University of Utah  
 Rm. 308, Biomedical Polymers Research Bldg., 20 S. 2030 E., SL, UT 84112, USA  
 Tel: 801 585 5606  
 Fax: 801 585 7177  
 Email: ddurn@genetics.utah.edu  
 Insert Length: 10000 Std Error: 0.00  
 Plate: 0199 row: D column: 23  
 Seq primer: CACACAGGAACAGCTATGACC  
 Class: plasmid ends  
 High quality sequence stop: 283.

FEATURES  
 source  
 1..283  
 /organism="Mus musculus"  
 /mol\_type="genomic DNA"  
 /strain="C57BL/6J"  
 /db\_xref="taxon:10090"  
 /clone="JUGC2M019P23"  
 /sex="female"  
 /lab\_host="E. coli strain XL10-Gold, Tl-resistant, F-"  
 /clone\_lib="Mouse 10kb plasmid JUGC2M library"  
 /notes="Vector: PWD42nv; Purified genomic DNA from M. musculus C57BL/6J (female) was obtained from the Jackson Laboratory Mouse DNA Resource  
 (http://www.jax.org/resources/documents/dnares/). The DNA was hydrodynamically sheared by repeated passage through a 0.005 inch orifice at constant velocity. The sheared DNA was blunt end-repaired with T4 DNA polymerase and T4 polynucleotide kinase. Adaptor oligonucleotides were ligated to the blunt ends in high molar excess. The adaptor DNA was purified and size-selected for a 9.5 to 10.5 kb range using preparative agarose gel electrophoresis. Vector DNA was prepared from a derivative of PWD42 (g|4732114|9b|AF129072.1), a copy-number inducible derivative of plasmid R1. The vector was ligated with adaptors complementary to the insert adaptors and purified. The sheared, adaptor mouse DNA was annealed to adaptor vector DNA, and transformed into chemically-competent E. coli XL10-Gold (Stratagene) cells and selected for ampicillin resistance."

ORIGIN  
 Query Match 80.9%; Score 17.8; DB 28; Length 283;  
 Best Local Similarity 90.5%; Pred. No. 1.7e+03;  
 Matches 19; Conservative 0; Mismatches 2; Indels 0; Gaps 0;  
 Qy 1 TGAGAGAGTTCTGGGTCTCT 21  
 Db 50 TGGAGAGAGTCTGGGTCTCT 70

RESULT 12  
 CC513479  
 LOCUS CH240\_358G5.T7 CHORI-240 Bos taurus genomic clone CH240\_358G5,  
 DEFINITION genomic survey sequence.  
 ACCESSION CC513479  
 VERSION CC513479.1 GI:31831767  
 KEYWORDS GSS.  
 SOURCE Bos taurus (cow)  
 ORGANISM Bos taurus  
 Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;

REFERENCE 1 (bases 1 to 296)  
 AUTHORS Holt,R., Stott,J., Yang,G., Barber,S., Smalls,D., Prabhu,A.-L., Tsai,M., Cloutier,A., Lee,D., Ghrn,N., Olson,T., Mayo,M., Butterfield,Y., Kirkpatrick,R., Liu,J., Guin,R., Chan,A., Chiu,R., Mathewson,C., Wye,N., Masson,A., Brown-John,M., Jones,S., Schein,J., Warra,M., de Jong,P., McWilliam,S., Barris,W., Dairymple,B.P. and Tellam,R.  
 TITLE Bovine BAC End Sequences from Library CHORI-240, PLATES 294 to 398  
 JOURNAL Unpublished (2003)  
 COMMENT Other GSSs: CH240\_358G5.TARBAC13P2  
 Contact: Rob Holt  
 Sequencing  
 The British Columbia Cancer Agency Genome Science Centre  
 600 W. 10th Ave, Vancouver, British Columbia, Canada V5Z 4S6  
 Tel: 604-877-6085  
 Fax: 604-877-6276  
 Email: rholt@bccsc.ca  
 Clones are derived from the bovine BAC library CHORI-240  
 (http://www.chori.org/bacpac/bovine240.htm). For BAC library availability, please contact Pieter de Jong (pdejong@mail.cho.org). Clones may be purchased from BACPAC Resources  
 (http://www.chori.org/bacpac/ordering-information.htm). This work was undertaken as part of the International Bovine BAC Mapping Consortium (IBBMC) by CSIRO Livestock Industries, Australia and the British Columbia Genome Sciences Centre, Canada.  
 Plate: 358 row: G column: 5  
 Seq primer: T7  
 Class: BAC ends.

FEATURES  
 Location/Qualifiers  
 1..296  
 /organism="Bos taurus"  
 /mol\_type="genomic DNA"  
 /strain="Breed: Hereford"  
 /db\_xref="taxon:9913"  
 /clone="CH240\_358G5"  
 /sex="Male"  
 /cell\_type="Blood"  
 /clone\_lib="CHORI-240"  
 /notes="Vector: PTARBAC1.3; Site 1: MboI; Site 2: MboI; Hereford bull Li Domino 99375; CHORI-240 Bovine BAC library (Male) produced by Pieter de Jong"

ORIGIN  
 Query Match 80.9%; Score 17.8; DB 29; Length 296;  
 Best Local Similarity 90.5%; Pred. No. 1.7e+03;  
 Matches 19; Conservative 0; Mismatches 2; Indels 0; Gaps 0;  
 Qy 1 TGAGAGAGTTCTGGGTCTCT 21  
 Db 19 TTAGAGAGTTCTGGGTCTCT 39

RESULT 13  
 BH834762  
 LOCUS BACPP7-C12.z Pristionchus pacificus BAC ends Pristionchus pacificus  
 DEFINITION genomic, genomic survey sequence.  
 ACCESSION BH834762  
 VERSION BH834762.1 GI:21036033  
 KEYWORDS GSS.  
 SOURCE Pristionchus pacificus  
 ORGANISM Pristionchus pacificus  
 Eukaryota; Metazoa; Nematoda; Chromadorea; Diplogasterida;  
 Neodiplogasteridae; Pristionchus.  
 REFERENCE 1 (bases 1 to 464)  
 AUTHORS Srinivasan,J., Sinz,W., Lanz,C., Brand,A., Nandakumar,R., Raddatz,G., Witte,H., Keller,H., Kipping,I., Pires da Silva,A., Jesse,T., Millare,J., de Both,M., Schuster,S.C. and Sommer,R.J.  
 TITLE A BAC-based genetic linkage map of the nematode Pristionchus pacificus  
 JOURNAL Unpublished (2002)

REFERENCE 1 (bases 1 to 296)  
 AUTHORS Holt,R., Stott,J., Yang,G., Barber,S., Smalls,D., Prabhu,A.-L., Tsai,M., Cloutier,A., Lee,D., Ghrn,N., Olson,T., Mayo,M., Butterfield,Y., Kirkpatrick,R., Liu,J., Guin,R., Chan,A., Chiu,R., Mathewson,C., Wye,N., Masson,A., Brown-John,M., Jones,S., Schein,J., Warra,M., de Jong,P., McWilliam,S., Barris,W., Dairymple,B.P. and Tellam,R.  
 TITLE Bovine BAC End Sequences from Library CHORI-240, PLATES 294 to 398  
 JOURNAL Unpublished (2003)  
 COMMENT Other GSSs: CH240\_358G5.TARBAC13P2  
 Contact: Rob Holt  
 Sequencing  
 The British Columbia Cancer Agency Genome Science Centre  
 600 W. 10th Ave, Vancouver, British Columbia, Canada V5Z 4S6  
 Tel: 604-877-6085  
 Fax: 604-877-6276  
 Email: rholt@bccsc.ca  
 Clones are derived from the bovine BAC library CHORI-240  
 (http://www.chori.org/bacpac/bovine240.htm). For BAC library availability, please contact Pieter de Jong (pdejong@mail.cho.org). Clones may be purchased from BACPAC Resources  
 (http://www.chori.org/bacpac/ordering-information.htm). This work was undertaken as part of the International Bovine BAC Mapping Consortium (IBBMC) by CSIRO Livestock Industries, Australia and the British Columbia Genome Sciences Centre, Canada.  
 Plate: 358 row: G column: 5  
 Seq primer: T7  
 Class: BAC ends.

FEATURES  
 Location/Qualifiers  
 1..296  
 /organism="Bos taurus"  
 /mol\_type="genomic DNA"  
 /strain="Breed: Hereford"  
 /db\_xref="taxon:9913"  
 /clone="CH240\_358G5"  
 /sex="Male"  
 /cell\_type="Blood"  
 /clone\_lib="CHORI-240"  
 /notes="Vector: PTARBAC1.3; Site 1: MboI; Site 2: MboI; Hereford bull Li Domino 99375; CHORI-240 Bovine BAC library (Male) produced by Pieter de Jong"

ORIGIN  
 Query Match 80.9%; Score 17.8; DB 29; Length 296;  
 Best Local Similarity 90.5%; Pred. No. 1.7e+03;  
 Matches 19; Conservative 0; Mismatches 2; Indels 0; Gaps 0;  
 Qy 1 TGAGAGAGTTCTGGGTCTCT 21  
 Db 19 TTAGAGAGTTCTGGGTCTCT 39

RESULT 13  
 BH834762  
 LOCUS BACPP7-C12.z Pristionchus pacificus BAC ends Pristionchus pacificus  
 DEFINITION genomic, genomic survey sequence.  
 ACCESSION BH834762  
 VERSION BH834762.1 GI:21036033  
 KEYWORDS GSS.  
 SOURCE Pristionchus pacificus  
 ORGANISM Pristionchus pacificus  
 Eukaryota; Metazoa; Nematoda; Chromadorea; Diplogasterida;  
 Neodiplogasteridae; Pristionchus.  
 REFERENCE 1 (bases 1 to 464)  
 AUTHORS Srinivasan,J., Sinz,W., Lanz,C., Brand,A., Nandakumar,R., Raddatz,G., Witte,H., Keller,H., Kipping,I., Pires da Silva,A., Jesse,T., Millare,J., de Both,M., Schuster,S.C. and Sommer,R.J.  
 TITLE A BAC-based genetic linkage map of the nematode Pristionchus pacificus  
 JOURNAL Unpublished (2002)

```

COMMENT      Contact: Sommer RJ
              Evolutionary Biology
              Max-Planck-Institute for Developmental Biology
              Spemannstr. 37-39, Tuebingen D-72076, Germany
              Tel: 00497071601371
              Fax: 00497071601498
              Email: ralf.sommer@tuebingen.mpg.de
              Class: BAC ends.
              Location/Qualifiers
FEATURES     source
              1..464
              /organism="Pristionchus pacificus"
              /mol_type="genomic DNA"
              /strain="var. California"
              /db_xref="taxon:54126"
              /clone_lib="Pristionchus pacificus BAC ends"
ORIGIN
Query Match      80.9%; Score 17.8; DB 28; Length 464;
Best Local Similarity 90.5%; Pred. No. 1.9e+03;
Matches 19; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
QY 1 TGAGAGAGTTCTGGGTGCTCT 21
Db 315 TGAGAGAGTTATGGCTGTCTCT 335

RESULT 14
CF606764
LOCUS          498 bp mRNA linear EST 30-SEP-2003
DEFINITION    GEMMAC1_000158 Grape Bud pSPORT1 Library Vitis vinifera cDNA 5',
              mRNA sequence.
ACCESSION     CF606764
VERSION       CF606764.1 GI:37187411
KEYWORDS      EST.
SOURCE        Vitis vinifera
ORGANISM      Vitis vinifera
              Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
              Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
              rosids; Vitaceae; Vitis.
REFERENCE     1 (bases 1 to 498)
AUTHORS      Moser,C., Segala,C., Fontana,P., Salakhudinov,I., Gatto,P.,
              Pindo,M., Zyprian,E., Toepfer,R., Grando,M.S. and Velasco,R.
              Expressed sequence tags from different organs of Vitis vinifera
              Unpublished (2003)
JOURNAL
COMMENT      Contact: Moser C
              Laboratorio di Genetica Molecolare
              Istituto Agrario di San Michele all'Adige (IASMA)
              via E. Mach 1, San Michele all'Adige (TN), I-38010, Italia
              Tel: 0039-0461-615314
              Fax: 0039-0461-650956
              Email: claudio.moser@ismaa.it
              The sequencing work has been funded by the 'Fondazione Cassa di
              Risparmio di Trento e Rovereto'
              High quality sequence stop: 498.
FEATURES     source
              1..498
              Location/Qualifiers
              /organism="Vitis vinifera"
              /mol_type="mRNA"
              /cultivar="Pinot noir"
              /db_xref="taxon:29760"
              /sex="Hermafrodite"
              /dev_stage="bud swelling"
              /lab_host="DH103"
              /clone_lib="Grape Bud pSPORT1 Library"
              /note="Organ: Bud; Vector: pSPORT1; Site_1: NotI; Site_2:
              SalI"
ORIGIN
Query Match      80.9%; Score 17.8; DB 14; Length 498;
Best Local Similarity 90.5%; Pred. No. 2e+03;
Matches 19; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
QY 2 GAGAGAGTTCTGGGTGCTCTA 22

COMMENT      Contact: Sommer RJ
              Evolutionary Biology
              Max-Planck-Institute for Developmental Biology
              Spemannstr. 37-39, Tuebingen D-72076, Germany
              Tel: 00497071601371
              Fax: 00497071601498
              Email: ralf.sommer@tuebingen.mpg.de
              Class: BAC ends.
              Location/Qualifiers
FEATURES     source
              1..601
              /organism="Pristionchus pacificus"
              /mol_type="genomic DNA"
              /strain="var. California"
              /db_xref="taxon:54126"
              /clone_lib="Pristionchus pacificus BAC ends"
ORIGIN
Query Match      80.9%; Score 17.8; DB 28; Length 601;
Best Local Similarity 90.5%; Pred. No. 2.1e+03;
Matches 19; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
QY 1 TCAGAGAGTTCTGGGTGCTCT 21
Db 315 TCAGAGAGTTATGGCTGTCTCT 335

Search completed: February 20, 2004, 01:41:44
Job time : 149.664 secs

```

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: February 19, 2004, 21:32:00 ; Search time 3.56994 Seconds  
(without alignments)  
3419.919 Million cell updates/sec

Title: US-09-939-853A-142

Perfect score: 22

Sequence: 1 TGAGAGAGTCTGGGTGCTCTA 22

Scoring table: IDENTITY NUC

Gapop 10.0 , Gapext 1.0

Searched: 682709 seqs, 277475446 residues

Total number of hits satisfying chosen parameters: 1365418

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Issued Patents NA.\*

1: /cgn2\_6/ptodata/2/ina/5A-COMB.seq.\*  
2: /cgn2\_6/ptodata/2/ina/5B-COMB.seq.\*  
3: /cgn2\_6/ptodata/2/ina/6A-COMB.seq.\*  
4: /cgn2\_6/ptodata/2/ina/6B-COMB.seq.\*  
5: /cgn2\_6/ptodata/2/ina/PCUTS-COMB.seq.\*  
6: /cgn2\_6/ptodata/2/ina/backfiles1.seq.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	17.2	78.2	1245	4	US-09-489-039A-3698
2	16.4	74.5	1273	3	US-08-725-758A-3
3	16.4	74.5	1373	3	US-08-725-758A-1
C 4	16.2	73.6	618	4	US-09-621-976-788
5	16.2	73.6	2713	2	US-08-916-901-6
6	16.2	73.6	2713	4	US-09-154-602-6
C 7	15.8	71.8	274	4	US-09-313-294A-5461
C 8	15.8	71.8	283	4	US-09-313-294A-4815
C 9	15.8	71.8	288	4	US-09-313-294A-809
C 10	15.8	71.8	288	4	US-09-313-294A-2911
11	15.8	71.8	835	4	US-09-833-381-1328
12	15.8	71.8	2081	4	US-09-853-839-7
13	15.8	71.8	2109	4	US-09-853-839-5
14	15.8	71.8	2172	4	US-09-853-839-3
15	15.8	71.8	2220	4	US-09-853-839-1
16	15.8	71.8	2353	4	US-09-622-880B-2
17	15.8	71.8	2806	4	US-09-653-839-9
18	15.8	71.8	3138	4	US-09-622-880B-16
C 19	15.6	70.9	291	4	US-09-313-294A-6747
20	15.6	70.9	311	4	US-08-621-976-11814
21	15.6	70.9	331	4	US-08-621-976-11984
22	15.6	70.9	344	4	US-09-621-976-12361
23	15.6	70.9	344	4	US-09-621-976-13537
24	15.6	70.9	7400	1	US-07-674-852-1
25	15.6	70.9	7400	1	US-08-473-185-1
26	15.6	70.9	7400	3	US-09-171-387-3
27	15.6	70.9	9573	4	US-09-220-132-168

28 15.4 70.0 2972 2 US-08-720-484A-3  
29 15.4 70.0 2972 3 US-08-953-823A-3  
30 15.4 70.0 2972 4 US-09-398-239-3  
31 15.4 70.0 2972 4 US-09-560-876A-3  
32 15.4 70.0 3301 4 US-09-148-545-66  
33 15.2 69.1 747 4 US-08-630-915A-39  
34 15.2 69.1 1167 4 US-09-679-686B-9  
35 15.2 69.1 1472 1 US-08-123-181A-9  
36 15.2 69.1 1472 1 US-08-483-278-9  
37 15.2 69.1 1558 4 US-09-123-030-7  
38 15.2 69.1 2389 3 US-08-691-563C-52  
39 15.2 69.1 2389 4 US-03-374-766-52  
40 15.2 69.1 2389 4 US-08-979-847B-48  
41 15.2 69.1 2464 4 US-09-620-312D-448  
42 15.2 69.1 2584 3 US-08-758-662-8  
43 15.2 69.1 2873 4 US-08-630-915A-193  
44 15.2 69.1 2905 4 US-09-595-684B-24  
45 15.2 69.1 3051 4 US-09-409-604-1

#### ALIGNMENTS

##### RESULT 1

US-09-489-039A-3698  
; Sequence 3698, Application US/09489039A  
; Patent No. 6610836

; GENERAL INFORMATION:

; APPLICANT: Gary Breton et. al

; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO KLEBSIELLA  
; FILE REFERENCE: 2709.2004001  
; CURRENT APPLICATION NUMBER: US/09/489,039A  
; CURRENT FILING DATE: 2000-01-27

; PRIOR APPLICATION NUMBER: US 60/117,747

; PRIOR FILING DATE: 1999-01-29

; NUMBER OF SEQ ID NOS: 14342

; SEQ ID NO 3698

; LENGTH: 1245

; TYPE: DNA

; ORGANISM: Klebsiella pneumoniae

US-09-489-039A-3698

Query Match 78.2%; Score 17.2; DB 4; Length 1245;  
Best Local Similarity 86.4%; Pred. No. 30;  
Matches 19; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

Oy 1 TGAGAGAGTCTGGGTGCTCTA 22  
Db 800 TGAGCGATTCTGGATGCTCTA 821

##### RESULT 2

US-08-725-758A-3

; Sequence 3, Application US/08725758A

; Patent No. 6160108

; GENERAL INFORMATION:

; APPLICANT: Reed, Guy

; APPLICANT: Clement, Christophe Y.

; TITLE OF INVENTION: NOVEL PLATELET ACTIVATION PROTEIN

; NUMBER OF SEQUENCES: 4

; CORRESPONDENCE ADDRESS:

; ADDRESSEE: Fish & Richardson P.C.

; STREET: 225 Franklin Street

; CITY: Boston

; STATE: MA

; COUNTRY: USA

; ZIP: 02110-2804

; COMPUTER READABLE FORM:

; MEDIUM TYPE: Diskette

; COMPUTER: IBM Compatible

; OPERATING SYSTEM: DOS

; SOFTWARE: FastSEQ Version 2.0

```

;
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/725,758A
; FILING DATE: 04-OCT-1996
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 60/005,074
; FILING DATE: 06-OCT-1995
; ATTORNEY/AGENT INFORMATION:
; NAME: Fraser, Janis K.
; REGISTRATION NUMBER: 34,819
; REFERENCE/DOCKET NUMBER: 05433/020001
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 617-542-5070
; TELEFAX: 617-542-8906
; TELEX: 200154
; INFORMATION FOR SEQ ID NO: 3:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1273 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: both
; TOPOLOGY: linear
; MOLECULE TYPE: cDNA
; FEATURE:
; NAME/KEY: Coding Sequence
; LOCATION: 44...1273
; US-08-725-758A-3

Query Match 74.5%; Score 16.4; DB 3; Length 1273;
Best Local Similarity 94.4%; Pred. No. 71;
Matches 17; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 TGAGAGAGTTCGGGTGT 18
Db 372 TGAGAGAGTTCGGGTGT 389

RESULT 3
US-08-725-758A-1
; Sequence 1, Application US/08725758A
; Patent No. 6160108
; GENERAL INFORMATION:
; APPLICANT: Reed, Guy
; APPLICANT: Clement, Christophe Y.
; TITLE OF INVENTION: NOVEL PLATELET ACTIVATION PROTEIN
; NUMBER OF SEQUENCES: 4
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Fish & Richardson P.C.
; STREET: 225 Franklin Street
; CITY: Boston
; STATE: MA
; COUNTRY: USA
; ZIP: 02110-2804
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; OPERATING SYSTEM: DOS
; SOFTWARE: FastSeq Version 2.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/725,758A
; FILING DATE: 04-OCT-1996
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 60/005,074
; FILING DATE: 06-OCT-1995
; ATTORNEY/AGENT INFORMATION:
; NAME: Fraser, Janis K.
; REGISTRATION NUMBER: 34,819
; REFERENCE/DOCKET NUMBER: 05433/020001
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 617-542-5070
; TELEFAX: 617-542-8906
; TELEX: 200154
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1373 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: both
; TOPOLOGY: linear
; MOLECULE TYPE: cDNA
; FEATURE:
; NAME/KEY: Coding Sequence
; LOCATION: 44...1321
; US-08-725-758A-1

Query Match 74.5%; Score 16.4; DB 3; Length 1373;
Best Local Similarity 94.4%; Pred. No. 72;
Matches 17; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 TGAGAGAGTTCGGGTGT 18
Db 372 TGAGAGAGTTCGGGTGT 389

RESULT 4
US-09-621-976-788/c
; Sequence 788, Application US/09621976
; Patent No. 6639063
; GENERAL INFORMATION:
; APPLICANT: Dumas Milne Edwards, J.B.
; APPLICANT: Jobert, S. J.Y.
; APPLICANT: Giordano, J.Y.
; TITLE OF INVENTION: ESTS and Encoded Human Proteins.
; FILE REFERENCE: GENSET.054PR2
; CURRENT APPLICATION NUMBER: US/09/621,976
; CURRENT FILING DATE: 2000-07-21
; NUMBER OF SEQ ID NOS: 19335
; SOFTWARE: Patent.pm
; SEQ ID NO 788
; LENGTH: 618
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 306...617
; US-09-621-976-788

Query Match 73.6%; Score 16.2; DB 4; Length 618;
Best Local Similarity 85.7%; Pred. No. 83;
Matches 18; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 TGAGAGAGTTCGGGTGTCT 21
Db 540 TCAGAGGGTTCGGGTGTCT 520

RESULT 5
US-08-916-901-6
; Sequence 6, Application US/08916901
; Patent No. 5892012
; GENERAL INFORMATION:
; APPLICANT: Hillman, Jennifer L.
; APPLICANT: Lal, Preeti
; APPLICANT: Corley, Neil C.
; APPLICANT: Shah, Purvi
; TITLE OF INVENTION: RAB PROTEINS
; NUMBER OF SEQUENCES: 9
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Incyte Pharmaceuticals, Inc.
; STREET: 3174 Porter Dr.
; CITY: Palo Alto
; STATE: CA
; COUNTRY: USA
; ZIP: 94304
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; OPERATING SYSTEM: DOS
; SOFTWARE: FastSeq for Windows Version 2.0

```



```
Db      259 TGATAGATTCTGGGTGGC 241
|||||
RESULT 9
US-09-313-294A-809/c
; Sequence 809, Application US/09313294A
; Patent No. 6476212
; GENERAL INFORMATION:
; APPLICANT: Lalgudi, Raghunath V.
; APPLICANT: Ito, Laura Y.
; APPLICANT: Sherman, Bradley K.
; TITLE OF INVENTION: POLYNUCLEOTIDES AND POLYPEPTIDES DERIVED FROM CORN EAR
; FILE REFERENCE: PL-0017 US
; CURRENT APPLICATION NUMBER: US/09/313,294A
; CURRENT FILING DATE: 1999-05-14
; NUMBER OF SEQ ID NOS: 7600
; SOFTWARE: PERL Program
; SEQ ID NO 809
; LENGTH: 288
; TYPE: DNA
; ORGANISM: Zea mays
; FEATURE:
; NAME/KEY: misc feature
; OTHER INFORMATION: Incyte ID No. 6476212 700549871H1
US-09-313-294A-809

Query Match      71.8%; Score 15.8; DB 4; Length 288;
Best Local Similarity 89.5%; Pred. No. 1.2e+02;
Matches 17; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY      1 TGAGAGATTCTGGGTGTC 19
|||||
Db      287 TGATAGATTCTGGGTGGC 269
|||||

RESULT 10
US-09-313-294A-2911/c
; Sequence 2911, Application US/09313294A
; Patent No. 6476212
; GENERAL INFORMATION:
; APPLICANT: Lalgudi, Raghunath V.
; APPLICANT: Ito, Laura Y.
; APPLICANT: Sherman, Bradley K.
; TITLE OF INVENTION: POLYNUCLEOTIDES AND POLYPEPTIDES DERIVED FROM CORN EAR
; FILE REFERENCE: PL-0017 US
; CURRENT APPLICATION NUMBER: US/09/313,294A
; CURRENT FILING DATE: 1999-05-14
; NUMBER OF SEQ ID NOS: 7600
; SOFTWARE: PERL Program
; SEQ ID NO 2911
; LENGTH: 288
; TYPE: DNA
; ORGANISM: Zea mays
; FEATURE:
; NAME/KEY: misc feature
; OTHER INFORMATION: Incyte ID No. 6476212 700553476H1
US-09-313-294A-2911

Query Match      71.8%; Score 15.8; DB 4; Length 288;
Best Local Similarity 89.5%; Pred. No. 1.2e+02;
Matches 17; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY      1 TGAGAGATTCTGGGTGTC 19
|||||
Db      260 TGATAGATTCTGGGTGGC 242
|||||

RESULT 11
US-09-833-381-1328
; Sequence 1328, Application US/09833381
; Patent No. 6672186
; GENERAL INFORMATION:
; APPLICANT: Robison, Keith E.
; TITLE OF INVENTION: No. 6672186el Nucleic Acid and Protein Homologs
; FILE REFERENCE: 5800-119
; CURRENT APPLICATION NUMBER: US/09/833,381
; CURRENT FILING DATE: 2001-04-11
; PRIOR APPLICATION NUMBER: 09/516,448
; PRIOR FILING DATE: 2000-02-29
; NUMBER OF SEQ ID NOS: 2050
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 1328
; LENGTH: 835
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-833-381-1328

Query Match      71.8%; Score 15.8; DB 4; Length 835;
Best Local Similarity 89.5%; Pred. No. 1.3e+02;
Matches 17; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY      3 AGACAGTTCTGGGTGTCCT 21
|||||
Db      755 AGACAGTTCTGGGTGTCCT 773
|||||

RESULT 12
US-09-653-839-7
; Sequence 7, Application US/09653839
; Patent No. 6433153
; GENERAL INFORMATION:
; APPLICANT: Donoho, Gregory
; APPLICANT: Turner, C. Alexander Jr.
; APPLICANT: Nehls, Michael
; APPLICANT: Friedrich, Glenn
; APPLICANT: Zambrowicz, Brian
; APPLICANT: Sands, Arthur T.
; TITLE OF INVENTION: No. 6433153el Human Calcium Dependent Proteases
; FILE REFERENCE: LEX-0038-USA
; CURRENT APPLICATION NUMBER: US/09/653,839
; CURRENT FILING DATE: 2000-09-01
; PRIOR APPLICATION NUMBER: US 60/152,057
; PRIOR FILING DATE: 1999-09-02
; NUMBER OF SEQ ID NOS: 9
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 7
; LENGTH: 2061
; TYPE: DNA
; ORGANISM: homo sapiens
US-09-653-839-7

Query Match      71.8%; Score 15.8; DB 4; Length 2061;
Best Local Similarity 89.5%; Pred. No. 1.4e+02;
Matches 17; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY      4 GAGAGTTCTGGGTGTCCTTA 22
|||||
Db      956 GGGAGTTCTGGGTGTCCTTA 974
|||||

RESULT 13
US-09-653-839-5
; Sequence 5, Application US/09653839
; Patent No. 6433153
; GENERAL INFORMATION:
; APPLICANT: Donoho, Gregory
; APPLICANT: Turner, C. Alexander Jr.
; APPLICANT: Nehls, Michael
; APPLICANT: Friedrich, Glenn
; APPLICANT: Zambrowicz, Brian
; APPLICANT: Sands, Arthur T.
; TITLE OF INVENTION: No. 6433153el Human Calcium Dependent Proteases
; FILE REFERENCE: LEX-0038-USA
```

FRICK AFFILIATION NUMBER: US 60/132,037



GenCore version 5.1.6  
Copyright (c) 1993 - 2004 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: February 19, 2004, 23:23:34 ; Search time 52.8809 Seconds  
(without alignments)  
1456.787 Million cell updates/sec

Title: US-09-939-853A-142

Perfect score: 22  
Sequence: 1 tsagagagttctgggtgctcta 22

Scoring table: IDENTITY NUC  
Gapop 10.0, Gapext 1.0

Searched: 2308684 seqs, 1750822206 residues

Total number of hits satisfying chosen parameters: 4617368

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : Published Applications NA:\*

1: /cgn2\_6/prodata/2/pubpna/US07\_PUBCOMB.seq:\*  
2: /cgn2\_6/prodata/2/pubpna/PCT\_NEW\_PUB.seq:\*  
3: /cgn2\_6/prodata/2/pubpna/US06\_NEW\_PUB.seq:\*  
4: /cgn2\_6/prodata/2/pubpna/US06\_PUBCOMB.seq:\*  
5: /cgn2\_6/prodata/2/pubpna/US07\_NEW\_PUB.seq:\*  
6: /cgn2\_6/prodata/2/pubpna/PCTUS\_PUBCOMB.seq:\*  
7: /cgn2\_6/prodata/2/pubpna/US08\_NEW\_PUB.seq:\*  
8: /cgn2\_6/prodata/2/pubpna/US08\_PUBCOMB.seq:\*  
9: /cgn2\_6/prodata/2/pubpna/US09\_PUBCOMB.seq:\*  
10: /cgn2\_6/prodata/2/pubpna/US09B\_PUBCOMB.seq:\*  
11: /cgn2\_6/prodata/2/pubpna/US09C\_PUBCOMB.seq:\*  
12: /cgn2\_6/prodata/2/pubpna/US09\_NEW\_PUB.seq:\*  
13: /cgn2\_6/prodata/2/pubpna/US10A\_PUBCOMB.seq:\*  
14: /cgn2\_6/prodata/2/pubpna/US10B\_PUBCOMB.seq:\*  
15: /cgn2\_6/prodata/2/pubpna/US10C\_PUBCOMB.seq:\*  
16: /cgn2\_6/prodata/2/pubpna/US10\_NEW\_PUB.seq:\*  
17: /cgn2\_6/prodata/2/pubpna/US60\_NEW\_PUB.seq:\*  
18: /cgn2\_6/prodata/2/pubpna/US60\_PUBCOMB.seq:\*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	% Match	Length	ID	Description
1	22	100.0	864	10	US-09-814-353-21302
2	20.4	92.7	763	9	Sequence 21302, A
3	17.8	80.9	2084	15	Sequence 953, App
4	17.4	79.1	2442	15	Sequence 1362, Ap
5	17.4	79.1	2466	15	Sequence 85, Appl
6	17.4	79.1	2789	15	Sequence 395, App
7	17.2	78.2	545	15	Sequence 22, Appl
8	17.2	78.2	641	15	Sequence 284738
9	17.2	78.2	904	15	Sequence 131511
10	17.2	78.2	1182	12	Sequence 131512
11	17.2	78.2	1182	12	Sequence 37180, A
12	17.2	78.2	1182	12	Sequence 39588, A
13	17.2	78.2	1185	12	Sequence 39447, A
14	16.8	76.4	2394	15	Sequence 23260, A
15	16.8	76.4	301692	15	Sequence 1043, Ap
					Sequence 11, Appl

15	16.4	74.5	1191	15	US-10-264-237-915	Sequence 915, App
17	16.4	74.5	1357	10	US-09-890-688-85	Sequence 85, Appl
18	16.4	74.5	1621	9	US-09-729-674-171	Sequence 171, App
19	16.4	74.5	2200	15	US-10-108-260A-1744	Sequence 1744, Ap
20	16.4	74.5	2560	12	US-10-424-599-130837	Sequence 130837,
21	16.4	74.5	3128	10	US-09-919-039-373	Sequence 373, App
22	16.4	74.5	3128	14	US-10-101-510-583	Sequence 583, Appl
23	16.4	74.5	27684	15	US-10-034-650-28	Sequence 28, Appl
24	16.2	73.6	179	14	US-10-029-386-17034	Sequence 17034, A
25	16.2	73.6	279	14	US-10-029-386-19880	Sequence 19880, A
26	16.2	73.6	294	9	US-09-864-761-19936	Sequence 19936, A
27	16.2	73.6	297	9	US-09-796-692-9356	Sequence 9356, Ap
28	16.2	73.6	297	14	US-10-040-862-9356	Sequence 9356, Ap
29	16.2	73.6	297	15	US-10-057-475B-9356	Sequence 9356, Ap
30	16.2	73.6	297	15	US-10-154-884B-9356	Sequence 9356, Ap
31	16.2	73.6	444	9	US-09-864-761-3157	Sequence 3157, Ap
32	16.2	73.6	457	15	US-10-027-632-56117	Sequence 56117, A
33	16.2	73.6	457	15	US-10-027-632-299308	Sequence 299308, A
34	16.2	73.6	502	15	US-10-027-632-61326	Sequence 61326, A2
35	16.2	73.6	525	14	US-10-029-386-20660	Sequence 20660, A6
36	16.2	73.6	542	10	US-09-951-936-1497	Sequence 1497, Ap
37	16.2	73.6	542	14	US-10-029-386-6149	Sequence 6149, Ap
38	16.2	73.6	555	14	US-10-029-386-22402	Sequence 22402, A
39	16.2	73.6	555	14	US-10-029-386-24654	Sequence 24654, A
40	16.2	73.6	593	14	US-10-029-386-3334	Sequence 3334, Ap
41	16.2	73.6	596	14	US-10-029-386-8674	Sequence 8674, Ap
42	16.2	73.6	596	14	US-10-029-386-10951	Sequence 10951, A
43	16.2	73.6	620	15	US-10-027-632-213371	Sequence 213371,
44	16.2	73.6	620	15	US-10-027-632-213372	Sequence 213372,
45	16.2	73.6	647	15	US-10-027-632-275671	Sequence 275671,

## ALIGNMENTS

RESULT 1  
US-09-814-353-21302  
; Sequence 21302, Application US/09814353  
; Publication No. US20030165831A1  
; GENERAL INFORMATION:  
; APPLICANT: Thompson, Pamela  
; APPLICANT: Lillie, James  
; TITLE OF INVENTION: NOVEL GENES, COMPOSITIONS, KITS, AND METHODS FOR  
; TITLE OF INVENTION: IDENTIFICATION, ASSESSMENT, PREVENTION, AND  
; TITLE OF INVENTION: THERAPY OF OVARIAN CANCER  
; FILE REFERENCE: MRI-006B  
; CURRENT APPLICATION NUMBER: US/09/814,353  
; CURRENT FILING DATE: 2001-03-21  
; PRIOR APPLICATION NUMBER: US 60/191,031  
; PRIOR FILING DATE: 2000-03-21  
; PRIOR APPLICATION NUMBER: US 60/207,124  
; PRIOR FILING DATE: 2000-05-25  
; PRIOR APPLICATION NUMBER: US 60/211,940  
; PRIOR FILING DATE: 2000-06-15  
; PRIOR APPLICATION NUMBER: US 60/216,820  
; PRIOR FILING DATE: 2000-07-07  
; PRIOR APPLICATION NUMBER: US 60/220,661  
; PRIOR FILING DATE: 2000-07-25  
; PRIOR APPLICATION NUMBER: US 60/257,672  
; PRIOR FILING DATE: 2000-12-21  
; NUMBER OF SEQ ID NOS: 22037  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 21302  
; LENGTH: 864  
; TYPE: DNA  
; ORGANISM: Homo sapiens  
; FEATURE:  
; NAME/KEY: misc feature  
; LOCATION: 1, 2, 3, 32, 862, 863, 864  
; OTHER INFORMATION: n = A,T,C or G  
US-09-814-353-21302

Query Match 100.0%; Score 22; DB 10; Length 864;  
Best Local Similarity 100.0%; Pred. No. 0.34; 0; Indels 0; Gaps 0;  
Matches 22; Conservative 0; Mismatches 0;

Qy 1 TGAGAGAGTCTGGGTGCTCTA 22  
Db 276 TGAGAGAGTCTGGGTGCTCTA 297

RESULT 2

US-09-867-550-953  
; Sequence 953, Application US/09867550  
; Patent No. US20020082206A1  
; GENERAL INFORMATION:  
; APPLICANT: Leach, Martin D.  
; APPLICANT: Mehraban, Fuad,  
; APPLICANT: Conley, Pamela  
; APPLICANT: Law, Debbie  
; APPLICANT: Topper, James  
; TITLE OF INVENTION: No. US20020082206A1 Polynucleotides from Atherogenic Cells and  
; FILE REFERENCE: 21402-013 (Cura-313)  
; CURRENT APPLICATION NUMBER: US/09/867,550  
; PRIOR FILING DATE: 2001-09-20  
; PRIOR APPLICATION NUMBER: USSN 60/208,427  
; PRIOR FILING DATE: 2000-05-30  
; NUMBER OF SEQ ID NOS: 2125  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 953  
; LENGTH: 763  
; TYPE: DNA  
; ORGANISM: Homo sapiens  
US-09-867-550-953

Query Match 92.7%; Score 20.4; DB 9; Length 763;  
Best Local Similarity 95.5%; Pred. No. 2.1;  
Matches 21; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 TGAGAGAGTCTGGGTGCTCTA 22  
Db 112 TGAGAGAGTCTGGGTGCTCTA 133

RESULT 3

US-10-108-260A-1362/c  
; Sequence 1362, Application US/10108260A  
; Publication No. US20040005560A1  
; GENERAL INFORMATION:  
; APPLICANT: HELIX RESEARCH INSTITUTE  
; TITLE OF INVENTION: No. US20040005560A1 full length cDNA  
; FILE REFERENCE: H1-A0106  
; CURRENT APPLICATION NUMBER: US/10/108,260A  
; PRIOR FILING DATE: 2002-03-27  
; NUMBER OF SEQ ID NOS: 5458  
; SOFTWARE: PatentIn Ver. 2.1  
; SEQ ID NO 1362  
; LENGTH: 2064  
; TYPE: DNA  
; ORGANISM: Homo sapiens  
US-10-108-260A-1362

Query Match 80.9%; Score 17.8; DB 15; Length 2064;  
Best Local Similarity 90.5%; Pred. No. 41;  
Matches 19; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 1 TGAGAGAGTCTGGGTGCTCT 21  
Db 159 TGAGAGAGTCTGGGTGCTCT 139

RESULT 4

US-10-080-334-85  
; Sequence 85, Application US/10080334

Publication No. US20040002584A1  
; GENERAL INFORMATION:  
; APPLICANT: Pena, Carol E. A.  
; APPLICANT: Shimkets, Richard A  
; APPLICANT: Li, Li  
; APPLICANT: Shenov, Suresh G  
; APPLICANT: Kekuda, Ramesh  
; APPLICANT: Spytek, Kimberly A  
; APPLICANT: Vernet, Corine A. M.  
; APPLICANT: Malyankar, Uriel M  
; APPLICANT: Guo, Xiaojia  
; APPLICANT: Gusev, Vladimir Y  
; APPLICANT: Casman, Stacie J  
; APPLICANT: Boldog, Ferenc L  
; APPLICANT: Furtak, Katarzyna  
; APPLICANT: Tchernev, Velizar T  
; APPLICANT: Patturajan, Meera  
; APPLICANT: Gangolli, Baha A  
; APPLICANT: Padigaru, Muralidhara  
; APPLICANT: Liu, Xiaohong  
; APPLICANT: Baumgartner, Jason C.  
; APPLICANT: Gerlach, Valerie  
; APPLICANT: Spaderna, Steven K  
; APPLICANT: Zehnusen, Bryan D  
; TITLE OF INVENTION: Proteins, Polynucleotides Encoding Them and Methods of  
; FILE REFERENCE: 21402-275  
; CURRENT APPLICATION NUMBER: US/10/080,334  
; CURRENT FILING DATE: 2002-02-21  
; PRIOR APPLICATION NUMBER: 60/270,523  
; PRIOR FILING DATE: 2001-02-21  
; PRIOR APPLICATION NUMBER: 60/322,712  
; PRIOR FILING DATE: 2001-09-17  
; PRIOR APPLICATION NUMBER: 60/311,980  
; PRIOR FILING DATE: 2001-08-13  
; PRIOR APPLICATION NUMBER: 60/330,307  
; PRIOR FILING DATE: 2001-10-18  
; PRIOR APPLICATION NUMBER: 60/278,796  
; PRIOR FILING DATE: 2001-03-26  
; PRIOR APPLICATION NUMBER: 60/281,521  
; PRIOR FILING DATE: 2001-04-04  
; PRIOR APPLICATION NUMBER: 60/276,677  
; PRIOR FILING DATE: 2001-03-16  
; PRIOR APPLICATION NUMBER: 60/311,595  
; PRIOR FILING DATE: 2001-08-10  
; PRIOR APPLICATION NUMBER: 60/270,220  
; PRIOR FILING DATE: 2001-02-21  
; PRIOR APPLICATION NUMBER: 60/274,295  
; PRIOR FILING DATE: 2001-03-08  
; PRIOR APPLICATION NUMBER: 60/318,526  
; PRIOR FILING DATE: 2001-09-10  
; PRIOR APPLICATION NUMBER: 60/286,548  
; PRIOR FILING DATE: 2001-04-25  
; PRIOR APPLICATION NUMBER: 60/291,765  
; PRIOR FILING DATE: 2001-05-17  
; PRIOR APPLICATION NUMBER: 60/270,797  
; PRIOR FILING DATE: 2001-02-23  
; PRIOR APPLICATION NUMBER: 60/276,400  
; PRIOR FILING DATE: 2001-03-16  
; PRIOR APPLICATION NUMBER: 60/270,810  
; PRIOR FILING DATE: 2001-02-23  
; NUMBER OF SEQ ID NOS: 388  
; SOFTWARE: PatentIn Ver. 2.1  
; SEQ ID NO 85  
; LENGTH: 2442  
; TYPE: DNA  
; ORGANISM: Homo sapiens  
US-10-080-334-85

Query Match 79.1%; Score 17.4; DB 15; Length 2442;  
Best Local Similarity 94.7%; Pred. No. 64;  
Matches 18; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

```

Qy 4 GAGAGTCTCGGTGTCCTA 22
    |||||
Db 1116 GAGAGTCTCGGTGTCCTA 1134

RESULT 5
US-10-159-563-396
; Sequence 396, Application US/10159563
; Publication No. US20040009154A1
; GENERAL INFORMATION:
; APPLICANT: Khan, Javed
; APPLICANT: Ringner, Markus
; APPLICANT: Peterson, Carsten
; APPLICANT: Meltzer, Paul
; TITLE OF INVENTION: SELECTIONS OF GENES AND METHODS OF USING THE SAME FOR
; TITLE OF INVENTION: DIAGNOSIS AND FOR TARGETING THE THERAPY OF SELECT CANCERS
; FILE REFERENCE: 11613 5SUS11
; CURRENT APPLICATION NUMBER: US/10/159,563
; CURRENT FILING DATE: 2002-12-09
; PRIOR APPLICATION NUMBER: US 10/133,937
; PRIOR FILING DATE: 2002-04-25
; NUMBER OF SEQ ID NOS: 444
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 396
; LENGTH: 2466
; TYPE: DNA
; ORGANISM: Homo sapiens
US-10-159-563-396

Query Match 79.1%; Score 17.4; DB 15; Length 2466;
Best Local Similarity 94.7%; Pred. No. 64;
Matches 18; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 4 GAGAGTCTCGGTGTCCTA 22
    |||||
Db 1184 GAGAGTCTCGGTGTCCTA 1202

RESULT 6
US-10-274-639-22
; Sequence 22, Application US/10274639
; Publication No. US20030232349A1
; GENERAL INFORMATION:
; APPLICANT: INCYTE GENOMICS, INC.
; APPLICANT: DELEGREANE, Angelo M.; GANDHI, Aneena R.
; APPLICANT: HAFALIA, April J.A.; LU, Dyung Aina M.
; APPLICANT: PATTERSON, Chandra; TRIBOULEY, Catherine M.
; APPLICANT: DAS, Debopriya; KALLICK, Deborah A.
; APPLICANT: NGUYEN, Daniel B.; LEE, Ernestine A.
; APPLICANT: KHAN, Farrah A.; YUE, Henry
; APPLICANT: AU-YOUNG, Janice K.; GRIFFIN, Jennifer A.
; APPLICANT: POLICKY, Jennifer L.; RAMKUNAR, Jayalaxmi
; APPLICANT: YANG, Junning; THANGAVELU, Kavitha
; APPLICANT: DING, Li; KEARNEY, Liam
; APPLICANT: BAUGHN, Mariah R.; BOROWSKY, Mark L.
; APPLICANT: SANJANWALA, Madhusudan M.; YAO, Monique G.
; APPLICANT: BURFORD, Neil; WALIA, Narinder K.
; APPLICANT: LAU, Preeti G.; LEE, Sally
; APPLICANT: TODD, Stephen; LO, Terence P.
; APPLICANT: TANG, Y. Tom; ELLIOTT, Vicki S.
; APPLICANT: AZIMZAI, Yalda; LU, Yan
; TITLE OF INVENTION: PROTEASES
; FILE REFERENCE: PI-0167 USA
; CURRENT APPLICATION NUMBER: US/10/274,639
; CURRENT FILING DATE: 2002-10-18
; PRIOR APPLICATION NUMBER: PCT/US01/22397
; PRIOR FILING DATE: 2001-07-17
; PRIOR APPLICATION NUMBER: US 60/220,063
; PRIOR FILING DATE: 2000-07-21
; PRIOR APPLICATION NUMBER: US 60/221,680
; PRIOR FILING DATE: 2000-07-28
; PRIOR APPLICATION NUMBER: US 60/223,544
; PRIOR FILING DATE: 2000-08-04

```

```

; PRIOR APPLICATION NUMBER: US 60/224,717
; PRIOR FILING DATE: 2000-08-11
; PRIOR APPLICATION NUMBER: US 60/225,988
; PRIOR FILING DATE: 2000-08-16
; PRIOR APPLICATION NUMBER: US 60/227,568
; PRIOR FILING DATE: 2000-08-23
; NUMBER OF SEQ ID NOS: 42
; SOFTWARE: PERL Program
; SEQ ID NO 22
; LENGTH: 2789
; TYPE: DNA
; ORGANISM: Homo sapiens
; NAME/KEY: misc feature
; OTHER INFORMATION: Incyte ID No. US20030232349A1 5155802CB1
US-10-274-639-22

Query Match 79.1%; Score 17.4; DB 15; Length 2789;
Best Local Similarity 94.7%; Pred. No. 64;
Matches 18; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 4 GAGAGTCTCGGTGTCCTA 22
    |||||
Db 1342 GAGAGTCTCGGTGTCCTA 1360

RESULT 7
US-10-027-632-284738
; Sequence 284738, Application US/10027632
; Publication No. US20030204075A9
; GENERAL INFORMATION:
; APPLICANT: Wang, David G.
; TITLE OF INVENTION: Identification and Mapping of Single Nucleotide
; TITLE OF INVENTION: Polymorphisms in the Human Genome
; FILE REFERENCE: 108827.129
; CURRENT APPLICATION NUMBER: US/10/027,632
; CURRENT FILING DATE: 2002-04-30
; PRIOR APPLICATION NUMBER: US 60/218,006
; PRIOR FILING DATE: 2000-07-12
; PRIOR APPLICATION NUMBER: US 60/198,676
; PRIOR FILING DATE: 2000-04-20
; PRIOR APPLICATION NUMBER: US 60/193,483
; PRIOR FILING DATE: 2000-03-29
; PRIOR APPLICATION NUMBER: US 60/185,218
; PRIOR FILING DATE: 2000-02-24
; PRIOR APPLICATION NUMBER: US 60/167,363
; PRIOR FILING DATE: 1999-11-23
; PRIOR APPLICATION NUMBER: US 60/156,358
; PRIOR FILING DATE: 1999-09-28
; PRIOR APPLICATION NUMBER: US 60/146,002
; PRIOR FILING DATE: 1999-08-09
; NUMBER OF SEQ ID NOS: 325720
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 284738
; LENGTH: 545
; TYPE: DNA
; ORGANISM: Human
US-10-027-632-284738

Query Match 78.2%; Score 17.2; DB 15; Length 545;
Best Local Similarity 86.4%; Pred. No. 79;
Matches 19; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

Qy 1 TGAGAGATTCTCGGTGTCCTA 22
    |||||
Db 196 TGAGACATTCTCGGTGTCATA 217

RESULT 8
US-10-027-632-131511/c
; Sequence 131511, Application US/10027632
; Publication No. US20030204075A9
; GENERAL INFORMATION:

```

; APPLICANT: Wang, David G.  
 ; TITLE OF INVENTION: Identification and Mapping of Single Nucleotide  
 ; POLYMORPHISMS IN THE HUMAN GENOME  
 ; FILE REFERENCE: 108827.129  
 ; CURRENT APPLICATION NUMBER: US/10/027,632  
 ; CURRENT FILING DATE: 2002-04-30  
 ; PRIOR FILING DATE: 2000-07-12 US 60/218,006  
 ; PRIOR APPLICATION NUMBER: US 60/198,676  
 ; PRIOR FILING DATE: 2000-04-20  
 ; PRIOR APPLICATION NUMBER: US 60/193,483  
 ; PRIOR FILING DATE: 2000-03-29  
 ; PRIOR APPLICATION NUMBER: US 60/185,218  
 ; PRIOR FILING DATE: 2000-02-24  
 ; PRIOR APPLICATION NUMBER: US 60/167,363  
 ; PRIOR FILING DATE: 1999-11-23  
 ; PRIOR APPLICATION NUMBER: US 60/156,358  
 ; PRIOR FILING DATE: 1999-09-28  
 ; PRIOR APPLICATION NUMBER: US 60/146,002  
 ; PRIOR FILING DATE: 1999-08-09  
 ; NUMBER OF SEQ ID NOS: 325720  
 ; SOFTWARE: FastSeq for Windows Version 4.0  
 ; SEQ ID NO 131511  
 ; LENGTH: 641  
 ; TYPE: DNA  
 ; ORGANISM: Human  
 US-10-027-632-131511

Query Match 78.2%; Score 17.2; DB 15; Length 641;  
 Best Local Similarity 86.4%; Pred. No. 80;  
 Matches 19; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 TGAGAGAGTTCTGGGTGTCCTA 22  
 DB 371 TGAGAGACTTCTGGTGACCCA 350

RESULT 9  
 US-10-027-632-131512  
 ; Sequence 131512, Application US/10027632  
 ; Publication No. US20030204075A9  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Wang, David G.  
 ; TITLE OF INVENTION: Identification and Mapping of Single Nucleotide  
 ; POLYMORPHISMS IN THE HUMAN GENOME  
 ; FILE REFERENCE: 108827.129  
 ; CURRENT APPLICATION NUMBER: US/10/027,632  
 ; CURRENT FILING DATE: 2002-04-30  
 ; PRIOR FILING DATE: 2000-07-12 US 60/218,006  
 ; PRIOR APPLICATION NUMBER: US 60/198,676  
 ; PRIOR FILING DATE: 2000-04-20  
 ; PRIOR APPLICATION NUMBER: US 60/193,483  
 ; PRIOR FILING DATE: 2000-03-29  
 ; PRIOR APPLICATION NUMBER: US 60/185,218  
 ; PRIOR FILING DATE: 2000-02-24  
 ; PRIOR APPLICATION NUMBER: US 60/167,363  
 ; PRIOR FILING DATE: 1999-11-23  
 ; PRIOR APPLICATION NUMBER: US 60/156,358  
 ; PRIOR FILING DATE: 1999-09-28  
 ; PRIOR APPLICATION NUMBER: US 60/146,002  
 ; PRIOR FILING DATE: 1999-08-09  
 ; NUMBER OF SEQ ID NOS: 325720  
 ; SOFTWARE: FastSeq for Windows Version 4.0  
 ; SEQ ID NO 131512  
 ; LENGTH: 904  
 ; TYPE: DNA  
 ; ORGANISM: Human  
 US-10-027-632-131512

Query Match 78.2%; Score 17.2; DB 15; Length 904;  
 Best Local Similarity 86.4%; Pred. No. 80;  
 Matches 19; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 TGAGAGAGTTCTGGGTGTCCTA 22  
 DB 272 TGAGAGACTTCTGGTGACCCA 293  
 RESULT 10  
 US-10-282-122A-37180  
 ; Sequence 37180, Application US/10282122A  
 ; Publication No. US20040029129A1  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Wang, Liangsu  
 ; APPLICANT: Zamudio, Carlos  
 ; APPLICANT: Malone, Cheryl  
 ; APPLICANT: Haselbeck, Robert  
 ; APPLICANT: Olsen, Kari  
 ; APPLICANT: Zyskind, Judith  
 ; APPLICANT: Wall, Daniel  
 ; APPLICANT: Trawick, John  
 ; APPLICANT: Carr, Grant  
 ; APPLICANT: Yamamoto, Robert  
 ; APPLICANT: Forsyth, R.  
 ; APPLICANT: Xu, H.  
 ; TITLE OF INVENTION: Identification of Essential Genes in Microorganisms  
 ; FILE REFERENCE: ELITRA.034A  
 ; CURRENT APPLICATION NUMBER: US/10/282,122A  
 ; CURRENT FILING DATE: 2003-02-20  
 ; PRIOR FILING DATE: 2000-03-21  
 ; PRIOR APPLICATION NUMBER: 60/191,078  
 ; PRIOR FILING DATE: 2000-03-21  
 ; PRIOR APPLICATION NUMBER: 60/206,848  
 ; PRIOR FILING DATE: 2000-05-23  
 ; PRIOR APPLICATION NUMBER: 60/207,727  
 ; PRIOR FILING DATE: 2000-05-26  
 ; PRIOR APPLICATION NUMBER: 60/230,335  
 ; PRIOR FILING DATE: 2000-09-06  
 ; PRIOR APPLICATION NUMBER: 60/230,347  
 ; PRIOR FILING DATE: 2000-09-09  
 ; PRIOR APPLICATION NUMBER: 60/242,578  
 ; PRIOR FILING DATE: 2000-10-23  
 ; PRIOR APPLICATION NUMBER: 60/253,625  
 ; PRIOR FILING DATE: 2000-11-27  
 ; PRIOR APPLICATION NUMBER: 60/257,931  
 ; PRIOR FILING DATE: 2000-12-22  
 ; PRIOR APPLICATION NUMBER: 60/267,636  
 ; PRIOR FILING DATE: 2001-02-09  
 ; PRIOR APPLICATION NUMBER: 60/269,308  
 ; PRIOR FILING DATE: 2001-02-16  
 ; Remaining Prior Application data removed - See File Wrapper or PALM.  
 ; NUMBER OF SEQ ID NOS: 78614  
 ; SOFTWARE: PatentIn version 3.1  
 ; SEQ ID NO 37180  
 ; LENGTH: 1163  
 ; TYPE: DNA  
 ; ORGANISM: Salmonella paratyphi A  
 US-10-282-122A-37180

Query Match 78.2%; Score 17.2; DB 12; Length 1163;  
 Best Local Similarity 86.4%; Pred. No. 80;  
 Matches 19; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 TGAGAGAGTTCTGGGTGTCCTA 22  
 DB 739 TGAGTGATTTCTGGATGTCCTA 760

RESULT 11  
 US-10-282-122A-38988  
 ; Sequence 38988, Application US/10282122A  
 ; Publication No. US20040029129A1  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Wang, Liangsu  
 ; APPLICANT: Zamudio, Carlos  
 ; APPLICANT: Malone, Cheryl

```

; APPLICANT: Haselbeck, Robert
; APPLICANT: Ohlsen, Kari
; APPLICANT: Zyskind, Judith
; APPLICANT: Wall, Daniel
; APPLICANT: Trawick, John
; APPLICANT: Carr, Grant
; APPLICANT: Yamamoto, Robert
; APPLICANT: Forsyth, R.
; APPLICANT: Xu, H.
; TITLE OF INVENTION: Identification of Essential Genes in Microorganisms
; FILE REFERENCE: ELITRA.034A
; CURRENT APPLICATION NUMBER: US/10/282,122A
; CURRENT FILING DATE: 2003-02-20
; PRIOR APPLICATION NUMBER: 60/191,078
; PRIOR FILING DATE: 2000-03-21
; PRIOR APPLICATION NUMBER: 60/206,848
; PRIOR FILING DATE: 2000-05-23
; PRIOR APPLICATION NUMBER: 60/207,727
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: 60/230,335
; PRIOR FILING DATE: 2000-09-06
; PRIOR APPLICATION NUMBER: 60/230,347
; PRIOR FILING DATE: 2000-09-09
; PRIOR APPLICATION NUMBER: 60/242,578
; PRIOR FILING DATE: 2000-10-23
; PRIOR APPLICATION NUMBER: 60/253,625
; PRIOR FILING DATE: 2000-11-27
; PRIOR APPLICATION NUMBER: 60/257,931
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: 60/267,636
; PRIOR FILING DATE: 2001-02-09
; PRIOR APPLICATION NUMBER: 60/269,308
; PRIOR FILING DATE: 2001-02-16
; Remaining Prior Application data removed - See File Wrapper or PALM.
; NUMBER OF SEQ ID NOS: 78614
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 38988
; LENGTH: 1182
; TYPE: DNA
; ORGANISM: Salmonella typhimurium
US-10-282-122A-38988

Query Match      78.2%; Score 17.2; DB 12; Length 1182;
Best Local Similarity 86.4%; Pred. No. 80;
Matches 19; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY      1 TGAGAGAGTTCTGGGTGTCCTA 22
        ||||| ||||| ||||| ||||| |||||
Db      740 TGAGTGATTCTGGATGTCCTA 761

RESULT 12
US-10-282-122A-39447
; Sequence 39447, Application US/10282122A
; Publication No. US20040029129A1
; GENERAL INFORMATION:
; APPLICANT: Wang, Liangsu
; APPLICANT: Zamudio, Carlos
; APPLICANT: Malone, Cheryl
; APPLICANT: Haselbeck, Robert
; APPLICANT: Ohlsen, Kari
; APPLICANT: Zyskind, Judith
; APPLICANT: Wall, Daniel
; APPLICANT: Trawick, John
; APPLICANT: Carr, Grant
; APPLICANT: Yamamoto, Robert
; APPLICANT: Forsyth, R.
; APPLICANT: Xu, H.
; TITLE OF INVENTION: Identification of Essential Genes in Microorganisms
; FILE REFERENCE: ELITRA.034A
; CURRENT APPLICATION NUMBER: US/10/282,122A
; CURRENT FILING DATE: 2003-02-20
; PRIOR APPLICATION NUMBER: 60/191,078
; PRIOR FILING DATE: 2000-03-21
; PRIOR APPLICATION NUMBER: 60/206,848
; PRIOR FILING DATE: 2000-05-23
; PRIOR APPLICATION NUMBER: 60/207,727
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: 60/230,335
; PRIOR FILING DATE: 2000-09-06
; PRIOR APPLICATION NUMBER: 60/230,347
; PRIOR FILING DATE: 2000-09-09
; PRIOR APPLICATION NUMBER: 60/242,578
; PRIOR FILING DATE: 2000-10-23
; PRIOR APPLICATION NUMBER: 60/253,625
; PRIOR FILING DATE: 2000-11-27
; PRIOR APPLICATION NUMBER: 60/257,931

```

```

; PRIOR FILING DATE: 2000-03-21
; PRIOR APPLICATION NUMBER: 60/206,848
; PRIOR FILING DATE: 2000-05-23
; PRIOR APPLICATION NUMBER: 60/207,727
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: 60/230,335
; PRIOR FILING DATE: 2000-09-06
; PRIOR APPLICATION NUMBER: 60/230,347
; PRIOR FILING DATE: 2000-09-09
; PRIOR APPLICATION NUMBER: 60/242,578
; PRIOR FILING DATE: 2000-10-23
; PRIOR APPLICATION NUMBER: 60/253,625
; PRIOR FILING DATE: 2000-11-27
; PRIOR APPLICATION NUMBER: 60/257,931
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: 60/267,636
; PRIOR FILING DATE: 2001-02-09
; PRIOR APPLICATION NUMBER: 60/269,308
; PRIOR FILING DATE: 2001-02-16
; Remaining Prior Application data removed - See File Wrapper or PALM.
; NUMBER OF SEQ ID NOS: 78614
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 39447
; LENGTH: 1182
; TYPE: DNA
; ORGANISM: Salmonella typhi
US-10-282-122A-39447

Query Match      78.2%; Score 17.2; DB 12; Length 1182;
Best Local Similarity 86.4%; Pred. No. 80;
Matches 19; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY      1 TGAGAGAGTTCTGGGTGTCCTA 22
        ||||| ||||| ||||| ||||| |||||
Db      740 TGAGTGATTCTGGATGTCCTA 761

RESULT 13
US-10-282-122A-23260
; Sequence 23260, Application US/10282122A
; Publication No. US20040029129A1
; GENERAL INFORMATION:
; APPLICANT: Wang, Liangsu
; APPLICANT: Zamudio, Carlos
; APPLICANT: Malone, Cheryl
; APPLICANT: Haselbeck, Robert
; APPLICANT: Ohlsen, Kari
; APPLICANT: Zyskind, Judith
; APPLICANT: Wall, Daniel
; APPLICANT: Trawick, John
; APPLICANT: Carr, Grant
; APPLICANT: Yamamoto, Robert
; APPLICANT: Forsyth, R.
; APPLICANT: Xu, H.
; TITLE OF INVENTION: Identification of Essential Genes in Microorganisms
; FILE REFERENCE: ELITRA.034A
; CURRENT APPLICATION NUMBER: US/10/282,122A
; CURRENT FILING DATE: 2003-02-20
; PRIOR APPLICATION NUMBER: 60/191,078
; PRIOR FILING DATE: 2000-03-21
; PRIOR APPLICATION NUMBER: 60/206,848
; PRIOR FILING DATE: 2000-05-23
; PRIOR APPLICATION NUMBER: 60/207,727
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: 60/230,335
; PRIOR FILING DATE: 2000-09-06
; PRIOR APPLICATION NUMBER: 60/230,347
; PRIOR FILING DATE: 2000-09-09
; PRIOR APPLICATION NUMBER: 60/242,578
; PRIOR FILING DATE: 2000-10-23
; PRIOR APPLICATION NUMBER: 60/253,625
; PRIOR FILING DATE: 2000-11-27
; PRIOR APPLICATION NUMBER: 60/257,931

```

; PRIOR FILING DATE: 2000-12-22  
 ; PRIOR APPLICATION NUMBER: 60/267,636  
 ; PRIOR FILING DATE: 2001-02-09  
 ; PRIOR APPLICATION NUMBER: 60/269,308  
 ; PRIOR FILING DATE: 2001-02-16  
 ; Remaining prior application data removed - See File Wrapper or PALM.  
 ; NUMBER OF SEQ ID NOS: 78614  
 ; SOFTWARE: PatentIn version 3.1  
 ; SEQ ID NO 23260  
 ; LENGTH: 1185  
 ; TYPE: DNA  
 ; ORGANISM: Klebsiella pneumoniae  
 US-10-282-122A-23260

Query Match 78.2%; Score 17.2; DB 12; Length 1185;  
 Best Local Similarity 86.4%; Pred. No. 80;  
 Matches 19; Conservative 0; Mismatches 3; Indels 0; Gaps 0;  
 QY 1 TGAGAGAGTTCTGGGTGTCCTA 22  
 |||||  
 Db 743 TGAGCGATTCTGGAGTGCCTA 764

RESULT 14  
 US-10-104-047-1043/c  
 ; Sequence 1043, Application US/10104047  
 ; Publication No. US20030236392A1  
 ; GENERAL INFORMATION:  
 ; APPLICANT: HELIX RESEARCH INSTITUTE  
 ; TITLE OF INVENTION: No. US20030236392A1 full length cdna  
 ; FILE REFERENCE: HI-A0105  
 ; CURRENT APPLICATION NUMBER: US/10/104,047  
 ; CURRENT FILING DATE: 2002-03-25  
 ; PRIOR APPLICATION NUMBER:  
 ; PRIOR FILING DATE:  
 ; NUMBER OF SEQ ID NOS: 4096  
 ; SOFTWARE: PatentIn Ver. 2.1  
 ; SEQ ID NO 1043  
 ; LENGTH: 2394  
 ; TYPE: DNA  
 ; ORGANISM: Homo sapiens  
 US-10-104-047-1043

Query Match 76.4%; Score 16.8; DB 15; Length 2394;  
 Best Local Similarity 90.0%; Pred. No. 1.3e+02;  
 Matches 18; Conservative 0; Mismatches 2; Indels 0; Gaps 0;  
 QY 2 GAGAGAGTTCTGGGTGTCCT 21  
 |||||  
 Db 1489 GAGAGATTCGAGTGTCAT 1470

RESULT 15  
 US-10-428-487-11/c  
 ; Sequence 11, Application US/10428487  
 ; Publication No. US20040006780A1  
 ; GENERAL INFORMATION:  
 ; APPLICANT: RASTELLI, LUCA K.  
 ; APPLICANT: GERBER, HANS-PETER  
 ; TITLE OF INVENTION: VEGF-MODULATED GENES AND METHODS EMPLOYING THEM  
 ; FILE REFERENCE: 09800080-0103  
 ; CURRENT APPLICATION NUMBER: US/10/428,487  
 ; CURRENT FILING DATE: 2003-05-02  
 ; PRIOR APPLICATION NUMBER: 09/815,153  
 ; PRIOR FILING DATE: 2001-03-21  
 ; PRIOR APPLICATION NUMBER: 60/191,201  
 ; PRIOR FILING DATE: 2000-03-22  
 ; NUMBER OF SEQ ID NOS: 84  
 ; SOFTWARE: PatentIn Ver. 2.1  
 ; SEQ ID NO 11  
 ; LENGTH: 301692  
 ; TYPE: DNA  
 ; ORGANISM: Homo sapiens

US-10-428-487-11  
 Query Match 76.4%; Score 16.8; DB 15; Length 301692;  
 Best Local Similarity 90.0%; Pred. No. 1.3e+02;  
 Matches 18; Conservative 0; Mismatches 2; Indels 0; Gaps 0;  
 QY 3 AGAGAGTTCTGGGTGTCCTA 22  
 |||||  
 Db 69276 AGAGATTGCTGGGTGTCCTA 69257  
 Search completed: February 20, 2004, 05:59:27  
 Job time : 59.8809 secs

